


```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Description

Result No.	Score	Query		DB	ID	Description
		Match	Length			
C 1	120	12.0	972	152	BG320800	BG320800 ZM04_02CC
C 2	116.4	11.7	543	166	BE355380	BE355380 DGL_114_E
C 3	106.6	10.7	489	166	BE355278	BE355278 DGL_114_E
C 4	106.6	10.7	562	122	AM924405	AM924405 WSL_69_BH
C 5	104.2	10.4	728	164	BE216453	BE216453 HY_CEB003
C 6	101.6	10.2	550	166	BE356583	BE356583 DGL_126_E
C 7	100.2	10.0	840	116	AM448084	AM448084 BRY_162A
C 8	99.6	10.0	453	155	BG556798	BG556798 EML_38_GG
C 9	99	9.9	549	154	BG644289	BG644289 EML_71_H
C 10	95.6	9.6	558	146	BE253449	BE253449 HYSMEF000
C 11	93.2	9.3	522	119	AM677520	AM677520 DGL_8_E02
C 12	93	9.3	629	140	BE775466	BE775466 MY_01-C-0-C
C 13	83	8.3	596	136	BE494103	BE494103 WHE1277_H
C 14	82.4	8.2	675	116	AM447980	AM447980 BRY_1163
C 15	82	8.2	486	174	BG158402	BG158402 FMI_60_AT
C 16	81.4	8.1	837	116	AM448031	AM448031 BRY_1270
C 17	80.2	8.0	407	167	BE406021	BE406021 WHE0405_
C 18	79.8	8.0	533	105	AL507042	AL507042 AL507042
C 19	79.8	8.0	700	105	AL576647	AL506647 AL506647
C 20	78.2	7.8	577	146	BE252720	BE252720 WHE2202_A
C 21	78.2	7.8	1338	76	AM731160	AM731160 GA_EA001
C 22	77.2	7.7	932	129	CNS00720	AL066742 Drosophila
C 23	75.4	7.5	796	162	BE041016	BE041016 OP16D06_C
C 24	75	7.5	574	140	BE804075	BE804075 sr82H01_Y
C 25	75	7.5	1427	138	BE636682	BE636682 rockefeller
C 26	74.8	7.5	1844	170	BE665160	BE665160 963057E093
C 27	74.4	7.4	614	140	BE803551	BE803551 sr60C06_Y
C 28	74.2	7.4	455	155	BG558793	BG558793 RH12_56
C 29	73.6	7.3	1307	76	AM731210	AM731210 GA_EA001
C 30	73.4	7.3	1725	170	BF660491	BF660491 963016G050
C 31	73	7.3	545	118	AM596826	AM596826 s183P10_Y
C 32	73	7.3	932	219	CNS00720	AL066742 Drosophila
C 33	72.8	7.3	1287	170	BE666361	BE666361 963067H093
C 34	72.6	7.3	456	155	BG560163	BG560163 RH12_71
C 35	72.4	7.2	591	166	BE357516	BE357516 DGL_20_GA
C 36	72.4	7.2	1158	76	AM727550	AM727550 GA_EA001
C 37	72.4	7.2	1338	76	AM731160	AM731160 GA_EA001
C 38	71.8	7.2	925	219	CNS0081P	AL053013 Drosophila
C 39	71.6	7.2	893	239	AZ193531	AZ193531 SP_1023_E
C 40	71.2	7.1	483	152	BG365747	BG365747 HYSME1000
C 41	71	7.1	925	219	CNS0091P	AL053013 Drosophila
C 42	71	7.1	1220	170	BF681909	BF681909 963030B013
C 43	70.6	7.1	1158	76	AM727550	AM727550 GA_EA001
C 44	70.2	7.0	634	228	AQ447725	AQ447725 mgx000066
C 45	70.2	7.0	935	219	CNS006XK	AL066051 Drosophila

ALIGNMENTS

RESULT	1
LOCUS	BG320800/c
DEFINITION	BG320800. 972 bp mRNA EST 27-FEB-2001
ACCESSION	Zm04_02c05_R Zm04_MAFc_ECORC_cold_stressed_malze_seedlings Zea mays
VERSION	BG320800
KEYWORDS	BG320800.1 GI:13150478
SOURCE	EST.
ORGANISM	Zea mays.
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea. 1 (bases 1 to 972)
AUTHORS	Singh,J.A., Waki,K., Couroux,P., De Moors,A., Harris,L.J., Hattori, .I., Ouellet,F., Robert,L.S., Sprout,D. and Tinker,N.A.
TITLE	Expressed Sequence Tags from Cold-Stressed Maize Seedlings

Unpublished (2001)
Contact: Singh, J. A.

Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-food Canada
960 Carling Avenue, Bldg. 20, Ottawa, Ontario, K1A 0C6, Canada
Tel: (613) 759-1662
Fax: (613) 759-1701
Email: singhja@agr.ca

Location/Qualifiers
1. .972

```

/organism="Zea mays"
/cultivar="CO328"
/db_xref="taxon:4577"
/clone="Zm04.02c05"
/clone_1b="Zm04.AAFC_ECORC_cold_stressed_maize_seedlings
/tissue_type="leaf, crown"
/notes="Vector: Bluescript SK-/XhoI-EcoRI. Site_1: Eco RI;
Site_2: Xho I; lower temperature 50 C / hour from 22 to
120C; bring to 50 in 1 hour from 120C. leave at 50c 2 days
, photoperiod 16 hours. light intensity was 125 uE-1.
library prepared by in vivo mass excision from amplified
library."

```

188 a	326 c	272 g	181 t	5 others
-------	-------	-------	-------	----------

Query Match	12.0%	Score 120;	DB 152;	Length 972;
Best Local Similarity	55.1%	Pred. No. 1.8e-13;		
Matches 283;	Conservative 1;	Mismatches 216;	Indels 14;	Gaps 2

[illegible]

LOCUS	BE355380	543 bp	mRNA	EST	20-JUL-2000
DEFINITION	Dg1_114_B02.b2_A002 Dark Grown 1 (Dg1) Sorghum bicolor cDNA, mRNA sequence.				
ACCESSION	BE355380				

[illegible]

QY	919	gttacc---gtgtccagaccagccttctgtgtgagcagcgcgagacgcgcgagcgcac	975
Db	74	GGTACCACGACGACCGCGTACGAGAGCTGCTCAACCGGAGACCCGACCGGCCCCAGGTCTT	15
QY	976	gtgc	979
Db	14	GTAC	11
RESULT	3		
LOCUS	BE355278	489 bp	EST
DEFINITION	DB1.114.B02.b1_A002	Dark Grown 1 (DG1)	Sorghum bicolor CDNA, mRNA
ACCESSION	BE355278		
VERSION	BE355278.1	GI:9296380	
KEYWORDS	EST.		
SOURCE	Sorghum.		
ORGANISM	Sorghum bicolor		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC clade; Panicoideae; Andropogoneae; Sorghum.		
AUTHORS	1 (bases 1 to 489) Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt, J.H.		
TITLE	An EST database from Sorghum: dark-grown seedlings		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: mpratt@uga.edu		
	Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.		
	Seq primer: JEN REV		
	High quality sequence stop: 483		
	PolyA-No.		
FEATURES	Location/Qualifiers		
SOURCE	1..489 /organism="Sorghum bicolor" /db_xref="taxon:4558" /clone_lib="Dark Grown 1 (DG1)" /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda zap; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A-RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."		
BASE COUNT	81 a 168 c 166 g 74 t		
ORIGIN			
Query Match	10.7%;	Score 106.6;	DB 166; Length 489;
Best Local Similarity	55.8%;	Pred. No. 6,3e-11;	
Matches 274; Conservative	0;	Mismatches 199;	Indels 18; Gaps 3
QY	492	cggcagatgcccagccgcgaagtctgacagcgcgcgagatagagacattgcgcctgagcga	551
Db	486	CGGTATATGTGCTGGCGGACCACTCGAGAGTGCCTCCAGTAAAGGCCATCCCTGCTGGA	427
QY	552	cgaagctgtccatgacctgcacacactcgtcccaagcgcgagaccggtatgtgtgagca	611
Db	426	TGACCAAGTTCATGGCGCGACCGCTCTCCTCGACGGGCGTGGAAGGCGTGGGGCGATGCG	367
QY	612	tctgttagaagtgcgtatgtgtgcagcgcgcagcctgtgcgcagatcccttcgcagagagcga	671
Db	366	AGTAGACGACGCTCGACCTAGTCCATGTCGAGGCGGCGGAGCGAGCGCGAGCCCTCGA	307
QY	672	tgaatgcccgcgcgcgaagccgcgcgtcttcgtttgacgcgcctcgtcatcttcgcgcgcacct	731
Db	306	CGATGTGCTTCCGGAGAGACCTTGTGCTGTGGGGCGCCGGGC---CGCCCCAGAAGACTT	250

ACCESSION	BG646289
VERSION	BG646289.1
KEYWORDS	EST.
SOURCE	Sorghum.
ORGANISM	Sorghum bicolor
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
TITLE	clade; Panicoideae; Andropogoneae; Sorghum.
JOURNAL	1 (bases 1 to 549)
COMMENT	Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H. An EST database from Sorghum: developing embryos Unpublished (2000) Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 542 1805 Email: emmpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Seq primer: JEN REV High quality sequence stop: 470 POLYA-No.
FEATURES	Location/Qualifiers
SOURCE	1..549 /organism="Sorghum bicolor" /db_xref="taxon:4558" /clone_lib="Embryo 1 (EM1)" /note="Organ: Embryos germinated for 24 hr; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from polyA RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT	92 a 189 c 182 g 85 t 1 others
ORIGIN	
Query Match	9.9%; Score 99; DB 154; Length 549;
Best Local Similarity	55.5%; Pred.No.1.7e-09;
Matches	261; Conservative 0; Mismatches 191; Indels 18; Gaps 3;
Dy	513 agttgcagacgagcagctatggagaacttcgctcgctgcagacgagcgtgtccatagcctgcc 572
Dd	548 ACTNAGAGGTGGCCCAAGTAAGGCCCATCTCGTGCATGACCAATTGATGCGCGCA 489
Oy	573 aaactcgtcccaacgagcagcagcgtcatgttgtagtcatcgytagaacgtcatgttgt 632
Dd	488 CGGTCTCGTCGACGCGGGCGTGAGGGGTGGGGCGATGGCAGTAAGACAGCTGAGAGTAGT 429
Oy	633 cgagcccaagcctgcgcagcagatcccctcgcaggaaggcagatgatgtgtccgcgcagacc 692
Dd	428 CCATGTCGACGCGGGCGGAGCGACCGCGGAGGCCCTCGCGCATGTGCTTCCGGGAGAGNC 369
Oy	693 cgcgtctgttcgagcgtcgcgtcatctcgcgcgcgcacaccttgctgcgcagacgagtgcct 752
Dd	368 CCTTTCTGTTGGGGCCCCGGG---CGCCCAGAAAGACTTGGTAGATGCCGAGAGTGG 312
Oy	753 cgcgcgttcgcgcgcgccttcgycgcagcacactgcacacaagcctcctcgtgtgtgtcccttgt 812
Dd	311 ACGCGCGCACGCCAGGTCGCCGAGTGCGCCTGCCCATGATCTCHCGGGGCGCCGCTT-- 252
Oy	813 agagccgcagcagcgttaactgttcgaggtgttcgagagcagttgatgtgcgcggtccggtcgt 872
Dd	253 -----GGCGTAGACCTTGGCGTTGCGAAGAAGTGAACGCCCGCCGATCCGSCACG 204
Oy	873 gttcatcagagcgcgcgcgtctgtcgtcctcgaacgcgtccgcgtgaaatgaac-----cgtgc 929
Dd	203 COTGAGACGACGCCCTTGGCTCTTCACAGTCCAGCTGCTGTGCCGAAGGTACACCACGAC 144
Oy	930 cgagccagagcctcgtgtgtgagcagcgcggaacgcgcagccgcagccgcacgtcgc 979

	RESULT	10	
Df	LOCUS	Bf253449/c	
	DEFINITION		
	ACCESSION	Bf253449	558 bp mRNA EST
	VERSION	Bf253449	HVMEF0001H06f Hordeum vulgare seedling root EST library HVCDNM0007
	KEYWORDS	(etioloated and unstressed) Hordeum vulgare cDNA clone	
	SOURCE	HVMEF0001H06f, mRNA sequence.	
	ORGANISM	Barley. Hordeum vulgare Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum. 1 (bases 1 to 558) Wing,R., Close,T.J., Kleinhoef,A., Wise,R., Begum,D., Frisch,D., Yu ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and Wood,T.. Development of a genetically and physically anchored EST resource for barley genomics Unpublished (2000) On Nov 16, 2000 this sequence version replaced gi:11182650. Contact: Wing RA Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Tel.: 864 656 7288 Fax: 864 656 4293 Email: rwing@clemson.edu Seq primer: AATTAACTCCACTCAATAAGG High quality sequence stop: 552. Location/qualifiers 1..558 /organism="Hordeum vulgare" /cultivar="Morex" /db_xref="taxon:4513" /clone="HVSMFE0001H06f" /clone_lib="Hordeum vulgare seedling root EST library HVCDNM0007 (etioloated and unstressed)" /tissue_type="Seedling root" /lab_host="TJC121" /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"	
	FEATURES		
	Source	BASE COUNT 105 a 197 c 179 g 73 t 4 others ORIGIN	
	Query Match	9.6%	Score 95.6; DB 146; Length 558;
	Best Local Similarity	55.2%;	Pred. No. 7.7e-09;
	Matches 254;	Conservative	0; Mismatches 188; Indels 18; Gaps 3;
Oy	513	aattgcagacagcggacttagagaccttgcgcgtcgagcacaggctgtccatgytcgcc	572
Dd	549	ACTCGNAGGTGGCCCCAAGRGAAGGCCAACGCCTGATCATGACCAGTTTCATGGCGCTTGA	490
Oy	573	acaactcglcccacagcgagccggaaccgcatgatgttgtatcatcbtgtaaacgttcagtlygt	632
Dd	489	CCTTGTCCTTCGATGGCGCTGGTGCGTGGGGCGGCTTGGCGGTAACAACGTCAGAGTTACT	430
Oy	633	cgaagcccaagctcgcagagcgatccctctgcagagaagcgatatgtatgccccgcgcgcagcgc	692
Dd	429	CCATTCAGAAGCGCGCTTAGAGAGAGCGCTTAGAGCGCTTCAGATGATGCTTGGCGGAGAGGC	370
Oy	693	cgccttcgttacgagcgcgcgtcatactcgcgcgcgcacacttgltgcgcacagcagtgatcct	752
Dd	369	CCTTGTCCTTGGGGCCCCCGGC---CGCCCAGAAAGAGCTTGSTGGAAGATGAAGACAGCTCGG	313
Oy	753	cgcgcgcgtcgcgcgccttggcgcagcacaactgtcccacacagctcctcggatgtgtgacctgt	812

Db	312	C6CGGCGCCAGGCAGGATGCCGGATGCTTGGCCCATGATTCCTCGGCGCGGCCGT--	253
OY	813	agagccgcgaagcgcgtacatgctggcggtgttcgaagcagtgtgatccgcggtgccggagcgt	872
Db	254	-----GGCGTAGACCTCGGCGCTTGCGAAGAATTGACC CGGCGCTCGCGCAGG	205
OY	873	ggtccatacaggcgagcgcgtgctgcctcagaacgcgtccgcctgaagttaac---cgtagc	929
Db	204	CCTGGAGGAGGAGACTTGGCTCTTTACGTCAAGCTGTGTTGGCGAAGGTATCACCAACGCGNC	145
OY	930	cgaaccagaagcctgctgtgtagcagcgcgcgaacccgcag	969
Db	144	CGTAGGACAGCTGGCTCACCCGAGCCCCGACCGGCGNCAG	105
RESULT 11			
LOCUS	AM677520/c		
DEFINITION	DG1.8.E02.bl_A002 Dark Grown 1 (DG1) Sorghum bicolor CDNA, mRNA		
ACCESSION	AM677520		
VERSION	AM677520.1		
KEYWORDS	EST.		
SOURCE	sorghum.		
ORGANISM	Sorghum bicolor		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum. 1 (bases 1 to 522) Cordonier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt, L.H.		
REFERENCE	An EST database from sorghum: dark-grown seedlings		
AUTHORS	Unpublished (2000)		
COMMENT	Contact: Cordonier-Pratt MM		

Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: JEN REV
High quality sequence stop: 465
POLYA-No.

FEATURES	source	Location/Qualifiers
		1..522
		/organism="Sorghum bicolor"
		/db_xref="taxon:4558"
		/clone_1ib="Dark grown 1 (DG1)"
		/note="Organ: 5-day-old dark-grown seedlings; Vector: lambda Zap; Site.1: XhoI; Site.2: EcoRI. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT		86 a 181 c 172 g 83 t
ORIGIN		

	Query Match	Similarity	9.3%	Score 93.2	DB 119	Length 522
	Best Local	Similarity	56.2%	Pred. No. 2.2e-08		
	Matches	245	Conservative	0	Mismatches 173	Indels 18
					Gaps	3
QY	547	ggcgcacgagcgtctcatatgctctgcacacccctctgtccacagcgcgagccgctcgatgt	606			
Db	522	gttcgatgacccagttcatatgcccgcacccctcttccttcgacggcggtgagagcgtcggggcg	463			
QY	607	gtgacatctcgtgtagagtcgatgtgtrtcgagcgccacgctcgtcgacagatattccctcgacga	666			
Db	462	atggcagtagtagacagacgcttcacgtagtccatgtrtcgaaaggcggcggagagccggagggcc	403			
QY	667	ggcgaatgtagtgcgcgcgcgcgacagcccggtgttcgttgagcgctgcgtcatcttcgcgcgc	726			

Db	402	CTCGACGATGTGCTTCCGGGAGAGACACCTGTGCTTGGGGGCCGGGG---CGCCCCAGAA	346
QY	727	gaacctggtgcgcagcaagtgctcctgcgcgcgtgcgcgccttgggcagccactgc	786
Db	345	GAGCTTGGTGGAGTACACACCTCGAGAGGGGGCCAGCCAGTCCGGATGGCTTGCCC	286
QY	787	caccagctcctcgtgtgtgcaccttgtagagccgcagccctacatctgcgcggtgtcgag	846
Db	285	CATGATTCCTCGGGGCGCCGCTT-----GGGCTAGACCTCGGCTTGTCGAA	238
QY	847	gcaattgatgcgcgcggtgcgcgcgcgtgtgtcactaagcgcgcgcggtgtgtcctctgac	906
Db	237	GAAATTGAAGCCCCGATCCCGGCACGCCCTGCAGCAGCGCTTGGCTCTTTCACGTCCAG	178
QY	907	gcctcgcctgaagttcac---cgtgcgcgcgcagagccctctgtgtgagcgcgcgaagc	963
Db	177	CTGCTTGGCCGAGTGCACCCACGACCGCTGACGAGAGCTGTCTACGCGGACCCCGACCG	118
QY	964	cccgagccgcgcgtgc	979
Db	117	GCCCAAGTTCTTGTAC	102

RESULT 12
 BE775466/c
 LOCUS
 DEFINITION MY-01-C-01 Pinfectansmy
 ACCESSION BE775466
 VERSION BE775466.1
 KEYWORDS GI:10229009
 SOURCE EST.
 ORGANISM potato late blight agent.
 Phytophthora infestans
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 Phytophthora.
 1 (bases 1 to 629)
 Kamoun,S., Hraber,P.T., Sobral,B.W.S., Nuss,D. and Govers,F.
 Initial assessment of gene diversity for the oomycete pathogen
 Phytophthora infestans based on expressed sequences
 Fungal Genet. Biol. 28 (2), 94-106 (1999)
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Govers F

FEATURES
 Laboratory of Phytopathology
 Wageningen University
 Binnenhaven 9, P.O. Box 8025, 6700 EE, Wageningen, The Netherlands
 Tel.: 31 317 483 138
 Fax: 31 317 483 412
 Email: Francine.Govers@medew.fyto.wau.nl.
SOURCE
 1. *Journal of Phytopathology* infected

```

/organism="Fungi"
/strain="DDR7602, Al mating type"
/db_xref="taxon:4787"
/clone_lib="PinfestansMY"
/dev_stage="4-week old vegetative, non-sporulating
mycelium in synthetic medium"
/lab_host="E. coli, strain DH5-alpha"
/ncbi_vector="pSPORT1; Site.1: SalI; Site.2: NotI; Total
RNA was isolated from mycelium of P. infestans DDR7602
cultured for 4 weeks in synthetic medium. EST clones were
named by their position in the microtiter plate, preceded
by the prefix MY (for mycelial) and the successive number
of the microtiter plate (e.g. MY-06-A-04)."
```

Query Match	9.3%	Score 93;	DB 140;	Length 629;
Best Local Similarity	47.5%	Pred. No. 2.4e-08;		
Matches 252;	Conservative 0;	Mismatches 279;	Indels 0;	Gaps 0;

Oy	425	aggtgtacagcactggttggaagacatgccacagagatgtgcgcgcgcgcgttctcc	484
Db	609	atggctactctgggtggtgttacacggcggcgcgacgaacgacccggcgccgacnrcg	550

[illegible]

LOCUS	BE494103	596 bp	mRNA	EST	02-AUG-2000
DEFINITION	WHE1277_H02_P03S	Secale cereale	anther cDNA	library	Secale cereale
ACCESSION	BE494103				
VERSION	BE494103				
KEYWORDS	EST.				
SOURCE	BE494103.1	GI:3660696			
ORGANISM	rye.				
REFERENCE	Secale cereale				
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Secale.				
TITLE	1 (bases 1 to 596)				
JOURNAL	Anderson,O.D., Butler,E., Chao,S., Choi,D.W., Close,T.J., Fenton				
COMMENT	R.D., Gustafson,J.P., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Ross,K., Seaton,C.L. and Tong,J.C.				
	The structure and function of the expressed portion of the wheat				
	genomes - Anther cDNA library from rye				
	Unpublished (2000)				
	Contact: Olin Anderson				
	US Department of Agriculture, Agriculture Research Service, Pacific				
	West Area, Western Regional Research Center				
	800 Buchanan Street, Albany, CA 94710, USA				
	Tel: 5105595773				
	Fax: 5105595818				
	Email: candersen@w.usda.gov				
	Sequence have been trimmed to remove vector sequence and low				
	quality sequence with phred score less than 20				
	Seq primer: Stratiagene SK primer.				
	Location/Qualifiers				
FEATURES	1..596				
source					

```

/organism="Secale cereale"
/cultivar="Blanco"
/db_xref="taxon:4550"
/clone="WHEH277_H02_P03"
/clone_lib="Secale cereale anther cDNA library"
/tissue_type="Anther"
/dev_stage="Adult plant before anthesis"

```

/lab host="E. coli SOLR"
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
 Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
 greenhouse. Anthers were harvested and pooled from early
 meiosis to late meiosis. The tissue, total RNA, and
 poly(A) RNA were prepared (Butler, Ross and Gustafson) at
 University of Missouri, Columbia. A cDNA library was
 made, and the cDNA clones were in vivo excised to give
 phagescript phagemids in the Tr Close lab (Choi, Close,
 Fenton) at the University of California, Riverside.
 Plasmid DNA preparations and DNA sequencing were performed
 in the OD Anderson lab (all other authors)."
 BASE COUNT 103 a 218 c 181 g 94 t
 ORIGIN

Query Match	8.3%	Score 83	DB 136	Length 596
Best Local Similarity	53.4%	Pred. No. 1.9e-06		
Matches 229	Conservative 0	Mismatches 185	Indels 15	Gaps 2

OY	515	ctgacgagcgagctagaggaaactctgcgcttgcgcagagcgctgtccatgcgctcgaac	574
Db	550	TCCGATTAACCCGAGTACTTACCTTCCCTCCCTCCACAAGCTTTTGAGCTACCCAAATC	491
OY	575	aactcgtcccaacggcgagcagcagctcgaatgtgtgcatactgttagacatcgaatgtgcg	634
Db	490	GTTGCTCTTGATGGGGATGGTGGTGTGCATGCGGGCTGTGATGAGAGGTGCATGTGATCG	431
OY	635	agcgccagcctgcgcgacgcatacctctgcgacgaggtgatagtgtgcgctgcgcgacgcccg	694
Db	430	ACGCGAGGCGGTCTCAGGCTGGCTCTCCACAGCAGCG-----GCCACGTACTCCGG	380
OY	695	ctgtcgttgaacgcgcctcgcctatctctgcgcgcgcagacctgtgtctgcacagcagtgctccgc	754
Db	379	CGCGCGCAGATGTGCTCTTCCCTGGGGCGTCCCGCTGGATCCGAACCTGTGTGGGCACC	320
OY	755	cgcgcgtccgcgcgcctcttggcgcagcaacctgtcccaacagctcctcgtgtgtgcctctgttag	814
Db	319	TGC-----ACCTGCTCTCCGCGGACGCTGCTTCAAGCCCTTGGCCCGAAGAGATCTGTTG	266
OY	815	agccgcagcgcgtacatcgtgcgaggtgtcgaagaaagttagatgcgcgcgtcccggtgcgtgcg	874
Db	265	GTTGTGGGCCCGCTTACCTCGCAAGGTGTCTGAAGAAGGTACGCCGCGCGGAAGGCGGTGC	206
OY	875	tccatcagcgcgacagcgcgtctgcgtctctcgcagcaggtctgcgtctgaagtctcacggtgcgcagc	934
Db	205	GTTGATGAGCGGAGCGCGCGCTCTGCTCCGACCGGGAGATTGTACGAGAGCTGTGTAGAGGCC	146
OY	935	cagagacgtcg 943	
Db	145	ATGCACCCG 137	

RESULT 14
AM447980/c
LOCUS AM447980 675 bp mRNA EST 03-JAN-2001
DEFINITION BR_1163 BRX Triticum aestivum CDNA clone P25-1D, mRNA sequence.
ACCESSION AM447980
VERSION AM447980.1 GI:12018515
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 675)
AUTHORS Clarke,B.C., Hobbs,M. and Appels,R.
TITLE Genes active in developing wheat endosperm
JOURNLS Unpublished (2000)
COMMENT Contact: Bryan Clarke
Division of Plant Industry
C.S.I.R.O.
GPO Box 1600, Canberra, ACT, Australia

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2001, 09:22:44 ; Search time 22.7 Seconds

(without alignments)
889.330 Million cell updates/sec

Title: US-09-463-705A-2

Perfect score: 1728

Sequence: 1 MTDDAATHVALGRSALLTSR.....LDELFPVAVSGAEPAMLO 333

Scoring table:

BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_0601.*
1: /SID8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1728	100.0	333	18 AAM19736	Sugar biosynthesis
2	1728	100.0	333	20 AAM99386	S.erythraea dtdp-4
3	1056	61.1	347	22 AAB61981	S. avermitilis ORF
4	418	24.2	348	20 AAY52837	Escherichia coli p
5	380.5	22.0	328	21 AAG39049	Arabidopsis thaliana
6	370	21.4	328	21 AAG35919	Arabidopsis thaliana
7	356	20.6	329	18 AAM23388	Zea mays protein f
8	353.5	20.5	412	21 AAG41675	Shaker-like potass
9	353.5	20.5	419	21 AAG41674	Arabidopsis thaliana
10	347.5	20.1	357	21 AAG41676	Arabidopsis thaliana
11	337	19.5	312	22 AAB79618	Cornebacterium gl

12	337	19.5	312	22 AAB80067	Cornebacterium gl
13	331	19.2	329	18 AAM23389	Shaker-like potass
14	329	19.0	344	19 AAM69711	Streptomyces clav
15	321	18.6	344	21 AAB08255	The 60S cell wall
16	317.5	18.4	340	21 AAG38206	Arabidopsis thaliana
17	317.5	18.4	349	21 AAG38205	Arabidopsis thaliana
18	317.5	18.4	351	21 AAG38204	Arabidopsis thaliana
19	301	17.4	337	21 AAG07306	Arabidopsis thaliana
20	301	17.4	346	21 AAG07305	Arabidopsis thaliana
21	301	17.4	349	21 AAG07304	Arabidopsis thaliana
22	300.5	17.3	345	21 AAG31466	Arabidopsis thaliana
23	299	17.3	366	21 AAG39050	Arabidopsis thaliana
24	285.5	16.5	319	21 AAG31467	Arabidopsis thaliana
25	284.5	16.5	266	21 AAG35920	Zea mays protein f
26	279.5	16.2	318	21 AAG40256	Arabidopsis thaliana
27	279.5	16.2	327	21 AAG40255	Arabidopsis thaliana
28	271.5	15.7	281	21 AAG54764	Arabidopsis thaliana
29	271	15.7	301	21 AAG40257	Arabidopsis thaliana
30	269	15.6	342	21 AAG29729	Arabidopsis thaliana
31	265	15.3	315	22 AAB76845	Cornebacterium gl
32	261.5	15.1	290	22 AAB76844	Cornebacterium gl
33	258.5	15.0	316	21 AAG29730	Arabidopsis thaliana
34	252	14.6	287	21 AAG30281	Arabidopsis thaliana
35	252	14.6	296	21 AAG30280	Arabidopsis thaliana
36	245.5	14.2	316	21 AAG35251	Zea mays protein f
37	245.5	14.2	371	21 AAG35250	Zea mays protein f
38	243.5	14.2	388	21 AAG35249	Zea mays protein f
39	243.5	14.1	270	21 AAG30282	Arabidopsis thaliana
40	242.5	14.0	375	21 AAG52050	Arabidopsis thaliana
41	242.5	14.0	377	21 AAG52049	Arabidopsis thaliana
42	242.5	14.0	386	21 AAG52048	Arabidopsis thaliana
43	237	13.7	284	21 AAG31468	Arabidopsis thaliana
44	234	13.5	316	21 AAG39500	Arabidopsis thaliana
45	234	13.5	324	21 AAG39499	Arabidopsis thaliana

ALIGNMENTS

RESULT 1
ID AAM19736 standard; Protein; 333 AA.
XX
AC AAM19736;
XX
DT 18-SEP-1997 (first entry)
XX
DE Sugar biosynthesis enzyme EryBII.
XX
KW Polyketide; glycosylation; eryBII; L-mycarose; antimicrobial;
KW antibiotic; antifungal; fungicide; anticancer; cytostatic;
KW antihelmintic.
XX
OS Saccharopolyspora erythraea.
XX
PN WO9723630-A2.
XX
PD 03-JUL-1997.
XX
PF 23-DEC-1996; 96WO-US20238.
XX
PR 21-DEC-1995; 95US-0576626.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Donadio S, Katz L, Staver MJ, Summers RG;
XX WPI. 1997-351066/32.
XX N-PSDB; AAT72684.
XX
PT New genes involved in sugar biosynthesis and attachment - used to
PT generate polyketide antimicrobials etc. with altered pattern of
PT glycosylation

XX Disclosure; Fig 4A; 85bp; English.
 PS EryBII (AAW19736) is an enzyme involved in the biosynthesis of the
 CC sugar L-mycarose. It is one of 10 enzymes (see also AAW19734-35 and
 CC AAW19737-43) predicted to be involved in D-desosamine or L-mycarose
 CC biosynthesis and attachment that have been identified from gene
 CC clusters (AAW72684 and AAW72685) of *Saccharopolyspora erythraea*. Novel
 CC glycosylation-modified polyketides are produced by selectively
 CC altering, inactivating or augmenting the *eryB* and/or *eryC* genes
 CC encoding these sugar biosynthesis enzymes and introducing them into
 CC polyketide-producing microorganisms.
 XX Sequence 333 AA:
 SQ

Query Match 100.0%; Score 1728; DB 18; Length 333;
 Best Local Similarity 100.0%; Pred. No. 4.5e-165;
 Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTTDAATHVRLGSRALLTSRLMLGTWNFSGRVEDDRLRLMDHARDGICNCLDTADMYGW 60
 1 mttdaathvrlgsrcalltsrlmlgtwnfsgrveddrlrlmdhardgicncltdadmygw 60
 Db 61 RLKKGTEELVGRMLAOGGGRREDTVLATKVGGMSESRVNDGSLARHITASCBSLRRL 120
 61 rlykghteelvgrmlaogggrrredtvlatkvggmseervndsglsarhitascbslrrl 120
 Db 121 GVPHIDVYOMHHIDRSAPMGEVMOAMDLSVASKGVSVGSSNFGWHIITAAOENAAARHS 180
 121 gvphidvygmhhidrsapmgevmoamdlsvasgkvsvvgssnfgwhiitaaoenaaarhs 180
 QY 181 LGMVSHOCLYNLAVRNALEVLPAQAQYGLGVFAMSPDHGGLSGALEKTLAAGTAVKSAQ 240
 181 lgmvshqclynlavrhaelevlpaagqyglgvfawspdhggllsgalektaagtavksaq 240
 Db 241 GRQAVLLPSLRPALEAVEKRCNLGEDPAEYGLAWLSRPGTAGAVIGPRTPEQDSALK 300
 241 grqavllpslrpaleavekrcnlgedpaeyglawlsrpgtagavigprtpeqldsalk 300
 Db 301 ASAMTIDEOALSELDEFPAVASGAAPPEAMLO 333
 301 asamtld eaglseldelfpavasgaapeawld 333
 QY

RESULT 2
 AAW9386
 ID AAW9386 standard; Protein; 333 AA.
 XX
 AC AAW9386;
 XX
 DT 08-JUN-1999 (first entry)
 XX
 DE *S. erythraea* dTDP-4-keto-L-6-deoxyhexose-2,3-reductase.
 XX
 KW Gene cluster; bacterium; enzyme; macrolide; antibiotic; erythromycin;
 KW secondary metabolite; eryBII; eryCIII; hybridisation; probe;
 KW glycosylation; macrolactone; oleandomycin.
 OS
 XX *Saccharopolyspora erythraea*.
 XX
 PN WO9905283-A2.
 XX
 PD 04-FEB-1999.
 XX
 PF 21-JUL-1998; 98WO-FR01593.
 XX
 PR 12-JUN-1998; 98FR-0007411.
 PR 25-JUL-1997; 97FR-0009458.
 XX
 PA (HMRI) HOECHST MARION ROUSSEL.
 XX
 PI Cortes J, Gaisser S, Leadlay P, Michel JM, Raynal MC;

PI Salah-Bey K, Fromentin C, Mendez C, Salas JA;
 XX
 DR WPI; 1999-142938/12.
 XX
 DR N-PSDB; AAX25772.
 XX
 PT New nucleic acid sequences encoding enzymes involved in macrolide
 PT biosynthesis - useful for producing hybrid secondary metabolites,
 PT particularly erythromycin analogues
 PS
 XX Claim 6; Fig 2; 221bp; French.
 CC This sequence represents the dTDP-4-keto-L-6-deoxyhexose-2,3-reductase
 CC enzyme encoded by the *eryBII* gene from the *eryE-eryAIII* gene cluster from
 CC the Gram-positive bacterium *Saccharopolyspora erythraea*. This gene
 CC cluster encodes enzymes involved in the production of the macrolide
 CC antibiotic erythromycin as a secondary metabolite. The erythromycin gene
 CC cluster spans approximately 53 kb and contains at least 20 open reading
 CC frames (ORF). The genes are used to produce hybrid secondary metabolites
 CC in *S. erythraea*, i.e. erythromycin analogues which may have improved
 CC properties or as hybridisation probes for isolating homologous genes
 CC involved in glycosylation of macrolactones in macrolide-producing
 CC strains (specifically oleandomycin-producing strains of *Streptomyces*
 CC antibioticus).
 XX
 SQ Sequence 333 AA:
 SQ

Query Match 100.0%; Score 1728; DB 20; Length 333;
 Best Local Similarity 100.0%; Pred. No. 4.5e-165;
 Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MTTDAATHVRLGSRALLTSRLMLGTWNFSGRVEDDRLRLMDHARDGICNCLDTADMYGW 60
 1 mttdaathvrlgsrcalltsrlmlgtwnfsgrveddrlrlmdhardgicncltdadmygw 60
 Db 61 RLKKGTEELVGRMLAOGGGRREDTVLATKVGGMSESRVNDGSLARHITASCBSLRRL 120
 61 rlykghteelvgrmlaogggrrredtvlatkvggmseervndsglsarhitascbslrrl 120
 Db 61 rlykghteelvgrmlaogggrrredtvlatkvggmseervndsglsarhitascbslrrl 120
 QY 121 GVPHIDVYOMHHIDRSAPMGEVMOAMDLSVASKGVSVGSSNFGWHIITAAOENAAARHS 180
 121 gvphidvygmhhidrsapmgevmoamdlsvasgkvsvvgssnfgwhiitaaoenaaarhs 180
 Db 121 gvphidvygmhhidrsapmgevmoamdlsvasgkvsvvgssnfgwhiitaaoenaaarhs 180
 QY 181 LGMVSHOCLYNLAVRNALEVLPAQAQYGLGVFAMSPDHGGLSGALEKTLAAGTAVKSAQ 240
 181 lgmvshqclynlavrhaelevlpaagqyglgvfawspdhggllsgalektaagtavksaq 240
 Db 241 GRQAVLLPSLRPALEAVEKRCNLGEDPAEYGLAWLSRPGTAGAVIGPRTPEQDSALK 300
 241 grqavllpslrpaleavekrcnlgedpaeyglawlsrpgtagavigprtpeqldsalk 300
 QY 301 ASAMTIDEOALSELDEFPAVASGAAPPEAMLO 333
 301 asamtld eaglseldelfpavasgaapeawld 333
 Db 301 asamtld eaglseldelfpavasgaapeawld 333
 QY

RESULT 3
 AAB61981
 ID AAB61981 standard; Protein; 347 AA.
 XX
 AC AAB61981;
 XX
 DT 14-MAY-2001 (first entry)
 XX
 DE *S. avermitilis* ORF8 (AvrI) polypeptide.
 XX
 KW Glycosylation; avermectin; aglycone; biocatalysis; antibiotic;
 KW oleandomycin; ORF8; AvrI.
 XX
 OS *Streptomyces avermitilis*.
 XX
 PN WO200109155-A1.
 XX

PD 08-FEB-2001.

XX 26-JUL-2000; 2000WO-US20331.

XX 30-JUL-1999; 99US-0146699.

XX (MERI) MERCK & CO INC.

XX MacNeil DJ, Occhi J, Gewain KM;

XX WPI; 2001-182929/18.

DR N-PSDB; AAC85191.

PT Novel nucleic acid fragments of Streptomyces avermitilis genome useful

PT for enzymatic, biochemical, biosynthetic and diagnostic purposes

XX Disclosure; Page 22; 63pp; English.

XX The invention relates to a 10 kb genomic DNA isolated from *S. avermitilis*

CC that contains genes encoding proteins for glycosylation of avermectin

CC aglycones. The polypeptides can be expressed by standard recombinant

CC methodology. The gene and the encoded polypeptides together with other

CC enzymatically active polypeptides, are useful to perform combinatorial

CC biocatalysis in vitro and in vivo in a host cell. They are useful for

CC performing biotransformations on macrocyclic compounds including avermectin

CC or other macrocyclic aglycones. The sequences are also useful in vivo in a

CC bacterial host, in vitro in combination with an actinomycete

CC fermentation, and in vitro in combination with an enzymatically active

CC polypeptides that are not from the avermectin biosynthetic pathway, to

CC effect the synthesis of a pharmaceutically active compound, e.g. an

CC antibiotic. Sequences AAB61973-981 represent polypeptides in the

CC avermectin biosynthetic pathway, involved in the synthesis and/or

CC addition of oleandrose to avermectin aglycones. The present sequence is

CC the *S. avermitilis* ORF8 (AVR1) polypeptide having the activity of

CC oleandrose synthesis.

XX Sequence 347 AA;

Query Match 61.1%; Score 1056; DB 22; Length 347;

Best Local Similarity 61.5%; Pred. No. 1.3e-97;

Matches 203; Conservative 35; Mismatches 92; Indels 0; Gaps 0;

YY 3 TDAATHVRLGNSALITSRMLCTVNFSGRVEDDRLRLMDARDGINCILDTADMYGWL 62

DB 17 ssaphtlhtlgrtlrlslalgtvnlgrveepearlmdhlaagtlldfclantgywrv 76

YY 63 YKGHEELVGRMLAOGGRRDVTATKVGEMSERVNDGSLARHTTASCEGSLRIGV 122

DB 77 hkyveevlygrwladrparregvialtkvqgdpmsgpnthglsvtnlvaacdaaltrltt 136

YY 123 DHIDVYQMHHDIDRSAPWDEVMQAMDSLVASGKVSYSVSSNFAQWHTIAAQAENARRHSIG 182

DB 137 dwidlyghldhrragvewwqamllltgkvrvygsnfagwlaasqaearrrnalg 196

YY 183 MWSHOCLYNLAVRAHELEYLPAQAYGICGVAFWSPLHGGLSGALKEKLAAGTAVYSAGQR 242

DB 197 laseqcyvnlvrhaelelpaasaygyvlywspbhgllygvlrktentavksagqr 256

YY 243 AQVILPSLRPAIEAVEKCRNLGDEPAEYGLAWLSRPIACAVIGPPELDDATKMS 302

DB 257 avealehhtltaayedvcadhjdpahvmaivstfpvgtvlgprteghvdalhal 316

YY 303 AMTLDEQALSELDEIFPAVAVSGAFAEWL 332

DB 317 rtlpelpvlarleellfpvgrgsapdawl 346

RESULT 4

AAV52837

ID AAV52837 standard; Protein: 348 AA.

XX AAV52837;

XX 26-JAN-2000 (first entry)

DT Escherichia coli protein sequence SEQ ID NO:4.

DE Isoprenoid: microorganism; detection; antibacterial; herbicide;

XX heart disease; osteoporosis; haemostasis; cancer; immunopotentialion;

XX heart food; antifouling coating; farnesyl pyrolinic acid; pyruvic acid;

XX 1-deoxy-D-xylulose-5-phosphate; glyceraldehyde-3-phosphate;

XX 2-C-methyl-D-erythreitol-4-phosphate.

XX Escherichia coli.

XX W09953071-A1.

XX 21-OCT-1999.

XX 14-APR-1999; 99WO-JP01987.

XX 14-APR-1998; 98JP-0103101.

XX 05-AUG-1998; 98JP-0221910.

XX 15-FEB-1999; 99JP-0035739.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Miyake K, Hashimoto S, Motoyama H, Ozaki A, Seto H, Kuzuyama T;

XX Takahashi S;

XX WPI; 1999-620434/53.

DR N-PSDB; AA233162.

PT Preparation of recombinant isoprenoid compounds useful for treatment of

PT heart diseases, osteoporosis and hemostasis, preventing cancer and

PT immunopotentialion

XX Claim 1; Page 70-72; 145pp; Japanese.

XX The present invention describes the preparation of an isoprenoid

CC compound comprising using at least 1 DNA e.g. encoding proteins which

CC elevate the efficiency of the synthesis or DNA encoding a farnesyl

CC pyrolinic producing enzyme. The method of preparation of an isoprenoid

CC compound comprises using at least 1 DNA, a vector, cloned cells, their

CC derived recombinant DNAs or transformed products in a culture system and

CC extracting the isoprenoid accumulated in the medium. The DNA encodes at

CC least 1 of the following: (a) a compound for activating or catalysing

CC the production of 1-deoxy-D-xylulose-5-phosphate from pyruvic acid and

CC glyceraldehyde-3-phosphate; (b) an enzyme producing farnesyl pyrolinic

CC acid; (c) a protein which elevates the efficiency of synthesis of

CC isoprenoid compounds and comprises a 3 or 4 amino acid sequence

CC optionally with 1 or more of the amino acids being deleted or

CC substituted or an additional amino acid being inserted; (d) a protein

CC which activates or catalyses the production of 2-C-methyl-D-erythreitol-

CC 4-phosphate from 1-deoxy-D-xylulose-5-phosphate; or (e) a protein which

CC activates a target compound or reaction and is a string end or hybrid of

CC the DNA encoded in (a)-(e). Isoprenoid compounds are useful in drugs

CC (e.g. for the treatment of heart diseases, osteoporosis and hemostasis,

CC for preventing cancer and as immunopotentialion), health foods and

CC antifouling coatings. The isoprenoids also inhibit enzymatic reactions

CC on the non-mevalonate pathway and can be used as antibacterials and

CC herbicides. The present sequence is used in the exemplification of

CC the present invention.

XX Sequence 348 AA;

Query Match 24.2%; Score 418; DB 20; Length 348;

Best Local Similarity 34.6%; Pred. No. 1.4e-33;

Matches 112; Conservative 52; Mismatches 124; Indels 36; Gaps 7;

YY 11 LGRSALITSRMLCTVNFSGRVEDDRLRLMDARDGINCILDTADMYGWL 62

DB 30 lgtklrlyrslcgtcmfgepdrghnawtlpessrplkrlagglntfdansys--- 86

```

0Y      63  KHNHEELUYGNWLAOGGRRREDYLAIRKVGSEMZERND--SGLSAHNHTIASEGSLRL 120
Db      87  -dgsseeivgral-rfdaircdvnavckv-----fhrygdipgegristqdlrfsldstql 140

0Y      121 GVDHLDVYQMHNIDRSAPFDEWMOAMDSTVASGKVSUYVSGSNFAGMWHIAAOCENAAARRHS 180
Db      141 gmdydlrlqrlrhwdrnprieetclalndvuykagrarlyagssmbhaagfaqlclgqhgw 200

0Y      181 LGMVSHOGLYMLAVNRHAELEVLVPAQAQVGLGVFMWSPRLHGGLLSG-----ALEKLA 221
Db      201 agfvmgmghyrllyreeeremrlpcyugvayavfwpvslargrlrlrpgettarlvsdewg 260

0Y      232 AGTAVKSAQGAQVLLPSLRPAIEAYEKFCNLTGEDPADEAGLAWMVSERPTACAVIGPRT 291
Db      261 knlykesendagf-----aerllygvseeigatraqvalawllskpylaapiilgstr 312

0Y      292 PEQDLDSALKASAMTLDQALSELD 315
Db      313 eeqldellnavdltltpqaeale 336

```

RESULT	5	
AG39049		
ID	AG39049	standard; Protein: 328 AA.
XX		
AC	AG39049;	
XX		
DT	18-OCT-2000	(first entry)
XX		
DE	Arabidopsis thaliana	protein fragment SEQ ID NO: 48263.
XX		
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.	
KW		
XX		
OS	Arabidopsis thaliana.	
XX		
PV	EP1033405-A2.	
XX		
PD	06-SEP-2000.	
XX		
PF	25-FEB-2000; 2000EP-0301439.	
XX		
PR	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127452.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	23-APR-1999;	99US-0130891.
PR	28-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	30-APR-1999;	99US-0132407.
PR	04-MAY-1999;	99US-0132484.
PR	05-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	07-MAY-1999;	99US-0132487.
PR	11-MAY-1999;	99US-0132863.
PR	14-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	18-MAY-1999;	99US-0134370.
PR	19-MAY-1999;	99US-0134768.
PR	20-MAY-1999;	99US-0134941.
PR		99US-0135124.

PR	21-MAY-1999	9905-0135353
PR	24-MAY-1999	9905-0135629
PR	25-MAY-1999	9905-0136021
PR	27-MAY-1999	9905-0136392
PR	28-MAY-1999	9905-0136782
PR	01-JUN-1999	9905-0137222
PR	03-JUN-1999	9905-0137528
PR	04-JUN-1999	9905-0137502
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138540
PR	10-JUN-1999	9905-0138847
PR	14-JUN-1999	9905-0139119
PR	16-JUN-1999	9905-0139452
PR	16-JUN-1999	9905-0139453
PR	17-JUN-1999	9905-0139462
PR	18-JUN-1999	9905-0139454
PR	18-JUN-1999	9905-0139455
PR	18-JUN-1999	9905-0139456
PR	18-JUN-1999	9905-0139457
PR	18-JUN-1999	9905-0139463
PR	18-JUN-1999	9905-0139750
PR	18-JUN-1999	9905-0139763
PR	21-JUN-1999	9905-0139817
PR	22-JUN-1999	9905-0139869
PR	23-JUN-1999	9905-0140053
PR	23-JUN-1999	9905-0140354
PR	24-JUN-1999	9905-0140625
PR	28-JUN-1999	9905-0140823
PR	29-JUN-1999	9905-0140391
PR	30-JUN-1999	9905-0141287
PR	01-JUL-1999	9905-0141842
PR	01-JUL-1999	9905-0142151
PR	02-JUL-1999	9905-0142055
PR	06-JUL-1999	9905-0142390
PR	08-JUL-1999	9905-0142803
PR	09-JUL-1999	9905-0143230
PR	12-JUL-1999	9905-0143277
PR	13-JUL-1999	9905-0143342
PR	14-JUL-1999	9905-0143624
PR	15-JUL-1999	9905-0144005
PR	16-JUL-1999	9905-0144085
PR	16-JUL-1999	9905-0144206
PR	19-JUL-1999	9905-0144325
PR	19-JUL-1999	9905-0144331
PR	19-JUL-1999	9905-0144332
PR	19-JUL-1999	9905-0144333
PR	19-JUL-1999	9905-0144334
PR	19-JUL-1999	9905-0144335
PR	20-JUL-1999	9905-0144352
PR	20-JUL-1999	9905-0144353
PR	20-JUL-1999	9905-0144354
PR	20-JUL-1999	9905-0144355
PR	20-JUL-1999	9905-0144356
PR	20-JUL-1999	9905-0144357
PR	20-JUL-1999	9905-0144358
PR	20-JUL-1999	9905-0144359
PR	20-JUL-1999	9905-0144360
PR	20-JUL-1999	9905-0144361
PR	20-JUL-1999	9905-0144362
PR	20-JUL-1999	9905-0144363
PR	20-JUL-1999	9905-0144364
PR	20-JUL-1999	9905-0144365
PR	20-JUL-1999	9905-0144366
PR	20-JUL-1999	9905-0144367
PR	20-JUL-1999	9905-0144368
PR	20-JUL-1999	9905-0144369
PR	20-JUL-1999	9905-0144370
PR	20-JUL-1999	9905-0144371
PR	20-JUL-1999	9905-0144372
PR	20-JUL-1999	9905-0144373
PR	20-JUL-1999	9905-0144374
PR	20-JUL-1999	9905-0144375
PR	20-JUL-1999	9905-0144376
PR	20-JUL-1999	9905-0144377
PR	20-JUL-1999	9905-0144378
PR	20-JUL-1999	9905-0144379
PR	20-JUL-1999	9905-0144380
PR	20-JUL-1999	9905-0144381
PR	20-JUL-1999	9905-0144382
PR	20-JUL-1999	9905-0144383
PR	20-JUL-1999	9905-0144384
PR	20-JUL-1999	9905-0144385
PR	20-JUL-1999	9905-0144386
PR	20-JUL-1999	9905-0144387
PR	20-JUL-1999	9905-0144388
PR	20-JUL-1999	9905-0144389
PR	20-JUL-1999	9905-0144390
PR	20-JUL-1999	9905-0144391
PR	20-JUL-1999	9905-0144392
PR	20-JUL-1999	9905-0144393
PR	20-JUL-1999	9905-0144394
PR	20-JUL-1999	9905-0144395
PR	20-JUL-1999	9905-0144396
PR	20-JUL-1999	9905-0144397
PR	20-JUL-1999	9905-0144398
PR	20-JUL-1999	9905-0144399
PR	20-JUL-1	

Query Match 22.0%; Score 380.5; DB 21; Length 328;
Best Local Similarity 30.6%; Pred. No. 7.3e-30;
Matches 102; Conservative 64; Mismatches 144; Indels 23; Gaps

DY 11 LGRSALLTSRLMGLT-VNFSGRVDDDALRMDIARPGICLDTADMYGKLYKGTFF 69
||| | | | | : : : : | : | : | : | : | : | :
Db 6 lqksqlkwtstlsfgawrlfngldvkeakslilgcrcrhgyffinaevya---ngraee 61

DY 70 LVGRMLAOGCGREDFTLVATKV--GGEMSERVNDSGTSARRHIIASCGSLRGLEVDHIDV 127
::: : : | | | | | : : : | : | : | : | : | : | :
Db 62 imggairelgrwrtsdvltstkftfyg---pgpnckjgtsrkhivgtktaskiklrmodydv 118

DY 128 YQHHHDISAPFDEWQAMDSLVASKVSIVGVSSNFAGMHIAAEENAAARRHSIGMVSHQ 187
| | | | | : : : | : | : | : | : | : | :
Db 119 lychnplastprieetvramuydlkgvaftygtsetwaagqtteawgadrdldvgpiweq 178

DY 188 CLYLAVLRH-AELEVLRAAQAYGCVFAMSPFHGGLLSGALEKIACGT---AVKSAGR 242
||: || | | | | : |: | | | | | : |: | : : :
Db 179 peymnfafhkveteflrplythtgiqltwspiasvltgkyknkaipdsdfalenynkl 238

DY 243 A-QVLLPSLRPAIEAYEKFCRNLEDPAEYGLAWLSRPRTAGAVIGPRPEQLDIALKA 301
| : | : | : | : | : | : | : | : | : | : | : | :
Db 239 anrlsvddvlrkvsqglrpdiadelvltlaqlaiawcaanpnvssvltgatresqlenmka 298

DY 302 SAMTLDEQALSE---LDEIFPVAVSGNAAPFAW 331
: | | | : | : | : | : | : | : | :
Db 299 ----vdvlpilltpivdkleqvlsqskprpesy 327

RESULT 6
AAG35919
ID AAG35919 standard; Protein: 328 AA.
XX
AC AAG35919:
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 43946.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter
termination sequence; corn.
OS Zea mays subsp. mays.
PN EPI033405-AZ.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 22.0%; Score 380.5; DB 21; Length 328;
Best Local Similarity 30.6%; Pred. No. 7.3e-30;
Matches 102; Conservative 64; Mismatches 144; Indels 23; Gaps

DY 11 LGRSALLTSRLMGLT-VNFSGRVDDDALRMDIARPGICLDTADMYGKLYKGTFF 69
||| | | | | : : : : | : | : | : | : | : | :
Db 6 lqksqlkwtstlsfgawrlfngldvkeakslilgcrcrhgyffinaevya---ngraee 61

DY 70 LVGRMLAOGCGREDFTLVATKV--GGEMSERVNDSGTSARRHIIASCGSLRGLEVDHIDV 127
::: : : | | | | | : : : | : | : | : | : | : | :
Db 62 imggairelgrwrtsdvltstkftfyg---pgpnckjgtsrkhivgtktaskiklrmodydv 118

DY 128 YQHHHDISAPFDEWQAMDSLVASKVSIVGVSSNFAGMHIAAEENAAARRHSIGMVSHQ 187
| | | | | : : : | : | : | : | : | : | :
Db 119 lychnplastprieetvramuydlkgvaftygtsetwaagqtteawgadrdldvgpiweq 178

DY 188 CLYLAVLRH-AELEVLRAAQAYGCVFAMSPFHGGLLSGALEKIACGT---AVKSAGR 242
||: || | | | | : |: | | | | | : |: | : : :
Db 179 peymnfafhkveteflrplythtgiqltwspiasvltgkyknkaipdsdfalenynkl 238

DY 243 A-QVLLPSLRPAIEAYEKFCRNLEDPAEYGLAWLSRPRTAGAVIGPRPEQLDIALKA 301
| : | : | : | : | : | : | : | : | : | : | : | :
Db 239 anrlsvddvlrkvsqglrpdiadelvltlaqlaiawcaanpnvssvltgatresqlenmka 298

DY 302 SAMTLDEQALSE---LDEIFPVAVSGNAAPFAW 331
: | | | : | : | : | : | : | : | :
Db 299 ----vdvlpilltpivdkleqvlsqskprpesy 327

RESULT 6
AAG35919
ID AAG35919 standard; Protein: 328 AA.
XX
AC AAG35919:
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 43946.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter
termination sequence; corn.
OS Zea mays subsp. mays.
PN EPI033405-AZ.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

PR	04-MAY-1999;	9905-0132484;	PR	22-JUL-1999;	9905-0145087;
PR	05-MAY-1999;	9905-0132485;	PR	22-JUL-1999;	9905-0145087;
PR	06-MAY-1999;	9905-0132486;	PR	22-JUL-1999;	9905-0145089;
PR	06-MAY-1999;	9905-0132487;	PR	22-JUL-1999;	9905-01451192;
PR	07-MAY-1999;	9905-0132488;	PR	22-JUL-1999;	9905-01451195;
PR	11-MAY-1999;	9905-0134256;	PR	23-JUL-1999;	9905-01452114;
PR	14-MAY-1999;	9905-0134218;	PR	23-JUL-1999;	9905-01452224;
PR	14-MAY-1999;	9905-0134219;	PR	26-JUL-1999;	9905-01452276;
PR	14-MAY-1999;	9905-0134221;	PR	27-JUL-1999;	9905-0145913;
PR	14-MAY-1999;	9905-0134370;	PR	27-JUL-1999;	9905-0145918;
PR	18-MAY-1999;	9905-0134768;	PR	27-JUL-1999;	9905-0145919;
PR	19-MAY-1999;	9905-0134941;	PR	28-JUL-1999;	9905-0145951;
PR	20-MAY-1999;	9905-0135121;	PR	02-AUG-1999;	9905-0146388;
PR	21-MAY-1999;	9905-0135353;	PR	02-AUG-1999;	9905-0146388;
PR	24-MAY-1999;	9905-0135629;	PR	02-AUG-1999;	9905-0146389;
PR	25-MAY-1999;	9905-0136021;	PR	03-AUG-1999;	9905-0147038;
PR	27-MAY-1999;	9905-0136392;	PR	04-AUG-1999;	9905-0147204;
PR	28-MAY-1999;	9905-0136782;	PR	04-AUG-1999;	9905-0147302;
PR	01-JUN-1999;	9905-0137222;	PR	05-AUG-1999;	9905-0147192;
PR	03-JUN-1999;	9905-0137528;	PR	05-AUG-1999;	9905-0147260;
PR	04-JUN-1999;	9905-0137502;	PR	06-AUG-1999;	9905-0147303;
PR	07-JUN-1999;	9905-0137724;	PR	06-AUG-1999;	9905-0147416;
PR	08-JUN-1999;	9905-0138094;	PR	09-AUG-1999;	9905-0147493;
PR	10-JUN-1999;	9905-0138540;	PR	09-AUG-1999;	9905-0147935;
PR	10-JUN-1999;	9905-0138847;	PR	10-AUG-1999;	9905-0148171;
PR	14-JUN-1999;	9905-0139119;	PR	11-AUG-1999;	9905-0148319;
PR	16-JUN-1999;	9905-0139452;	PR	12-AUG-1999;	9905-0148341;
PR	16-JUN-1999;	9905-0139453;	PR	13-AUG-1999;	9905-0148565;
PR	17-JUN-1999;	9905-0139492;	PR	13-AUG-1999;	9905-0148684;
PR	18-JUN-1999;	9905-0139454;	PR	16-AUG-1999;	9905-0149368;
PR	18-JUN-1999;	9905-0139455;	PR	17-AUG-1999;	9905-0148175;
PR	18-JUN-1999;	9905-0139456;	PR	18-AUG-1999;	9905-0149426;
PR	18-JUN-1999;	9905-0139457;	PR	20-AUG-1999;	9905-0149723;
PR	18-JUN-1999;	9905-0139458;	PR	20-AUG-1999;	9905-0149929;
PR	18-JUN-1999;	9905-0139459;	PR	23-AUG-1999;	9905-0149902;
PR	18-JUN-1999;	9905-0139460;	PR	23-AUG-1999;	9905-0149930;
PR	18-JUN-1999;	9905-0139461;	PR	23-AUG-1999;	9905-0150566;
PR	18-JUN-1999;	9905-0139462;	PR	26-AUG-1999;	9905-0150884;
PR	18-JUN-1999;	9905-0139463;	PR	27-AUG-1999;	9905-0151065;
PR	18-JUN-1999;	9905-0139750;	PR	27-AUG-1999;	9905-0151066;
PR	21-JUN-1999;	9905-0139763;	PR	27-AUG-1999;	9905-0151080;
PR	21-JUN-1999;	9905-0139817;	PR	30-AUG-1999;	9905-0151303;
PR	22-JUN-1999;	9905-0139889;	PR	31-AUG-1999;	9905-0151438;
PR	23-JUN-1999;	9905-0140353;	PR	01-SEP-1999;	9905-0151930;
PR	23-JUN-1999;	9905-0140354;	PR	01-SEP-1999;	9905-0152363;
PR	24-JUN-1999;	9905-0140693;	PR	07-SEP-1999;	9905-0153070;
PR	28-JUN-1999;	9905-0140823;	PR	10-SEP-1999;	9905-0153478;
PR	29-JUN-1999;	9905-0140991;	PR	13-SEP-1999;	9905-0154018;
PR	30-JUN-1999;	9905-0141287;	PR	15-SEP-1999;	9905-0154039;
PR	01-JUL-1999;	9905-0141842;	PR	16-SEP-1999;	9905-0154779;
PR	01-JUL-1999;	9905-0142154;	PR	20-SEP-1999;	9905-0155139;
PR	02-JUL-1999;	9905-0142055;	PR	22-SEP-1999;	9905-0155548;
PR	06-JUL-1999;	9905-0142			

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145813.
PR 27-JUL-1999; 99US-0145818.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147360.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148345.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

	Query Match	20.5%	Score 353.5:	DB: 21;	Length 412;
	Best Local Similarity	30.9%;	Pred. No.5.2e-27;		
	Matches 110;	Conservative	63;	Mismatches 140;	Indels 43; Gaps 11;
OY	2 TTDAATHVRLRSALTSRLMLGTNFSGSRVEDDALRLMDARGAINCLETDADMYGWR	61			
Db	52 tknameyrlkigdslntseveumgnttfgcgnhtekeshemlsylaiseeglnclidtaeaiprp	111			
OY	62 LYK---GHTTELVGWMLAQGGRRDDTYLAIRKVG--EKSERVNDSG---LSARHIITAS	112			
Db	112 mkkelqgktldiyisswllk-sqdrdklvtakvcgyserseyiridsgeilvrrdaanikes	169			
OY	113 CEGSLRRLGVNHIDPYVOHHHIDR-----SAPMEQVMQAMSLVASVKP	155			
Db	170 veketkrigtctydldlgphprdyprltfgdflyeteskwrpsvpfrteqratfdillivegkv	229			
OY	156 SYVGSSNFAGMHLAAOENNAARRHSLSGMVSHOCLYNLAVR--HAELEVLPAQAQYGLGV	212			
Db	230 ryilgvsnetsygtvetefvntakleglpkrivtsiqngyslllvrcryevdlvevchpkncnvgl	289			
OY	213 FAMSPLHGILGSLGALETALAAG-TAYKSAO-----GRAVLLPSL-RPIIEATKRCRNIG	265			
Db	290 layepblggsgstsg-kylatdqeatknarlmlfpyymeyksislakeatlqdvveavakkyyg	347			
OY	266 EDPAEVGLAWLVSRPGIAGAVIPPTPEQLDSALKASAMT--LD EQALSLELDITF	318			

Db	348	ltlpealglvrdpivstlllgatvskqlkexlidalimterpfsqgwmadlav	403
RESULT	9		
ID	AA641674	standard; Protein; 419 AA.	
XX	AA641674;		
XX			
DT	18-OCT-2000	(first entry)	
XX			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 51881.		
XX			
KW	Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter termination sequence.		
XX			
OS	Arabidopsis thaliana.		
XX			
PN	EP1033405-A2.		
XX			
ED	06-SEP-2000.		
XX			
PF	25-FEB-2000; 2000EP-0301439.		
XX			
PR	25-FEB-1999;	99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123548.	
PR	23-MAR-1999;	99US-0125788.	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130049.	
PR	23-APR-1999;	99US-0130510.	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048.	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132484.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132487.	
PR	07-MAY-1999;	99US-0132863.	
PR	11-MAY-1999;	99US-0134256.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	
PR	21-MAY-1999;	99US-0135353.	
PR	24-MAY-1999;	99US-0135629.	
PR	25-MAY-1999;	99US-0136021.	
PR	27-MAY-1999;	99US-0136392.	
PR	28-MAY-1999;	99US-0136782.	
PR	01-JUN-1999;	99US-0137222.	
PR	03-JUN-1999;	99US-0137528.	
PR	04-JUN-1999;	99US-0137502.	
PR	07-JUN-1999;	99US-0137724.	
PR	08-JUN-1999;	99US-0138094.	
PR	10-JUN-1999;	99US-0138540.	
PR	10-JUN-1999;	99US-0138847.	
PR	14-JUN-1999;	99US-0139119.	
PR	16-JUN-1999;	99US-0139452.	
PR	17-JUN-1999;	99US-0139453.	
PR	17-JUN-1999;	99US-0139452.	
PR	18-JUN-1999;	99US-0139454.	
PR	18-JUN-1999;	99US-0139455.	

PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147202.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.

PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152630.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 20.5%; Score 353.5; DB 21; Length 419;
Best Local Similarity 30.9%; Pred. No. 5, 3e-27;
Matches 110; Conservative 63; Mismatches 140; Indels 43; Gaps 11;

QY 2 TTPAATHVRLGRSALLTSRLTLGVNFSGRVEDDRLRLMDHARDGINCIDTADYGNR 61
DB 59 tknameyrligdsdlnisevtumtfcgeqntekeshemysaieegincidaaeypip 118
QY 62 LVR---GHTTELGRMLAOGGREDTVALTKVG--EMSERVNDG---LSARHIIAS 112
DB 119 mkketgkidllysswlk--sgqrkivlatkvcysetsayirsdgellrvdaanlkes 176
QY 113 CEGSLRRLGVHDHIDVYOMHHIDR-----SAPWDEVWQAMDSLVASGV 155

Db 177 vekskrtgtyidldlqhmprtypligdfyetskwrpsvpfiegqtrafdllivekv 236
Oy 156 SYVGSSNFGWHIAAQAENARRHSLGMVSHOCLYNLAVR--HAEEVLPPAAQAYGLGV 212
Db 237 ryigsnetsygvtfvntakleglpklvislqngysllivrcryevdlvevchpknenvgl 296
Oy 213 FAMSPLHGGILSGALEKTLAAG-TAVKSAO-----GRAQVLPLSL-RLPALEATKRCRNLG 265
Db 297 laysplggsgslsg--kylatdqeaclknarlnlfpymerykyslakeatigvvevakkyy 354
Oy 266 EDPAAVGLAWLSPRGIAAGAVIPRTPEQLDSALKASAMT--LDEQALSELDEIF 318
Db 355 ltpvelalgvtrdrpfvstligatsvqklkedidaflmterpfsgewmadidavf 410
RESULT 10
AAG41676
ID AAG41676 standard; Protein: 357 AA.
XX AAG41676;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 51883.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hydrialsation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EPI033405-A2.
PN 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
PE
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145227.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.

PR	05-ANG-1999	9905-01471992
PR	05-ANG-1999	9905-01472602
PR	06-ANG-1999	9905-0147303
PR	06-ANG-1999	9905-0147416
PR	09-ANG-1999	9905-0147493
PR	09-ANG-1999	9905-0147535
PR	10-ANG-1999	9905-0148171
PR	11-ANG-1999	9905-0148319
PR	12-ANG-1999	9905-0148341
PR	13-ANG-1999	9905-0148365
PR	13-ANG-1999	9905-0148684
PR	16-ANG-1999	9905-0149368
PR	17-ANG-1999	9905-0149175
PR	18-ANG-1999	9905-0149126
PR	20-ANG-1999	9905-0149122
PR	20-ANG-1999	9905-0149123
PR	20-ANG-1999	9905-0149292
PR	23-ANG-1999	9905-0149902
PR	23-ANG-1999	9905-0149930
PR	25-ANG-1999	9905-0150566
PR	26-ANG-1999	9905-0150884
PR	27-ANG-1999	9905-0151065
PR	27-ANG-1999	9905-0151066
PR	27-ANG-1999	9905-0151080
PR	30-ANG-1999	9905-0151303
PR	31-ANG-1999	9905-0151438
PR	01-SEP-1999	9905-0151330
PR	07-SEP-1999	9905-0153263
PR	10-SEP-1999	9905-0153070
PR	13-SEP-1999	9905-0153718
PR	15-SEP-1999	9905-0154018
PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154779
PR	22-SEP-1999	9905-0155139
PR	23-SEP-1999	9905-0155486
PR	24-SEP-1999	9905-0155659
PR	28-SEP-1999	9905-0156458
PR	29-SEP-1999	9905-0156596
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157753
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0158923
PR	13-OCT-1999	9905-0159294
PR	14-OCT-1999	9905-0159295
PR	14-OCT-1999	9905-0159330
PR	14-OCT-1999	9905-0159637
PR	18-OCT-1999	9905-0159638
PR	18-OCT-1999	9905-0159584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160767
PR	21-OCT-1999	9905-0160815
PR	22-OCT-1999	9905-0160880
PR	22-OCT-1999	9905-0160981
PR	22-OCT-1999	9905-0160988
PR	22-OCT-1999	9905-0161361
PR	25-OCT-1999	9905-0161404
PR	25-OCT-1999	9905-0161405
PR	25-OCT-1999	9905-0161406
PR	26-OCT-1999	9905-0161359
PR	26-OCT-1999	9905-0161360
PR	26-OCT-1999	9905-0161361
PR	28-OCT-1999	9905-0161920
PR	28-OCT-1999	9905-0161932
PR	28-OCT-1999	9905-0161943
PR	29-OCT-1999	9905-0162142

[illegible]

PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX
XX (BADI) BASF AG.
XX
XX Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI; 2001-061975/07.
XX
XX N-PSDB: AAF71735.
XX
XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
XX
XX metabolism and oxidative phosphorylation protein for production or
XX
XX modulation of production of fine chemicals e.g. amino acids,
XX
XX carbohydrates or enzymes -
XX
XX
XX Claim 20; Page 1208-1209; 1246pp; English.
XX
XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
XX
XX metabolism and oxidative phosphorylation (SMP) proteins given in
XX
XX AAB79243 to AAB 79633 which are involved in carbon metabolism and
XX
XX energy production. The C. glutamicum SMP gene can be used in vectors
XX
XX (II) for expression in host cells and production or modulation of
XX
XX production of fine chemicals, such as, an organic acid, a proteinogenic
XX
XX or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
XX
XX a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
XX
XX acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
XX
XX cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
XX
XX (II) encoded by them are used for diagnosing the presence or activity of
XX
XX Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells
XX
XX containing them are used to map genomes of organisms related to
XX
XX C. glutamicum, identify and localise C. glutamicum sequences of interest,
XX
XX in evolutionary studies, in determining SMP protein regions required
XX
XX for function, in modulating SMP protein activity, in modulating the
XX
XX metabolism of sugars, and in modulating high-energy molecule production
XX
XX in a cell (I.e. ATP, NADPH).
XX
XX
XX Sequence 312 AA;
SQ

Query Match 19.5%; Score 337; DB 22; Length 312;
Best Local Similarity 30.3%; Pred. No. 1.6e-25;

Matches 96; Conservative 53; Mismatches 150; Indels 18; Gaps 6;

QY 11 IGRSALTSRLTGTVNFSGREDDALRLMDHARDGNCIDTADMYGWRLYKGTTEEL.70
DB 2 VSSGSLIWSLIGLSTGWSGTELEAGDIFKafinsgglltdvspny---ttgvaem 57
QY 71 VGRWLAOGGRRREDVLAQKVGGE---MSERVNDGSLARHIIASCEGSLRLQGDHID 126
DB 58 lgtmI-daevsrsavvissagvnpalplgrtvd---crrnllqglvtlralntdyd 113
QY 127 VYOMHHIDRSAPFWEVQAMDSLVAAGKYSYVSSNFGACWHIAA--AQENARRHSLGAV 184
DB 114 lwsygywdegltphveadtldyavrtgyrvgvsgwqlavtchaasnaasarpv 173
QY 185 SHQCLYINLAVRHALELVPAQAYGAGVPAWSPRLHGSLISGLEKLAAGTAATKASQGRAG 244
DB 174 vaqneqslIerraegeIIPatqhlvgfIagaplggvtlakyreiphsdaasgtrda 233
QY 245 VLLPST---RPAIAYEKFCRNIGEDPAEVLAWLSPRGTAGAVTGRPEQDLSALK 300
DB 234 evqsyldmrgtrIvdaIdaakgIgispaavtattwrdprgtavtvgartheqshlIk 293

QY 301 ASAMTLEQALSELDEI 317
DB 294 aevtlpIptqaldav 310
RESULT 12
AAB80067
ID AAB80067 standard; Protein; 312 AA.
XX
XX AAB80067;
XX
XX 30-APR-2001 (first entry)
XX
XX Corynebacterium glutamicum MP protein sequence SEQ ID NO:868.
XX
XX
XX Corynebacterium glutamicum; metabolic pathway protein; MP protein;
XX
XX fine chemical production; microorganism; organic acid; nucleoside;
XX
XX nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
XX
XX lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
XX
XX carbohydrate; aromatic compound; cofactor; polyketide; enzyme;
XX
XX Corynebacterium glutamicum.
XX
XX WO200100843-A2.
XX
XX
XX 04-JAN-2001.
XX
XX
XX 23-JUN-2000; 2000WO-IB00923.
XX
XX
XX 25-JUN-1999; 99US-0141031.
XX
XX 01-JUL-1999; 99DE-1030476.
XX
XX 02-JUL-1999; 99US-0142101.
XX
XX 08-JUL-1999; 99DE-1031415.
XX
XX 08-JUL-1999; 99DE-1031418.
XX
XX 08-JUL-1999; 99DE-1031419.
XX
XX 08-JUL-1999; 99DE-1031420.
XX
XX 08-JUL-1999; 99DE-1031424.
XX
XX 08-JUL-1999; 99DE-1031428.
XX
XX 08-JUL-1999; 99DE-1031434.
XX
XX 08-JUL-1999; 99DE-1031435.
XX
XX 08-JUL-1999; 99DE-1031443.
XX
XX 08-JUL-1999; 99DE-1031453.
XX
XX 08-JUL-1999; 99DE-1031457.
XX
XX 08-JUL-1999; 99DE-1031465.
XX
XX 08-JUL-1999; 99DE-1031478.
XX
XX 08-JUL-1999; 99DE-1031510.
XX
XX 08-JUL-1999; 99DE-1031541.
XX
XX 08-JUL-1999; 99DE-1031573.
XX
XX 08-JUL-1999; 99DE-1031592.
XX
XX 08-JUL-1999; 99DE-1031632.
XX
XX 08-JUL-1999; 99DE-1031634.
XX
XX 08-JUL-1999; 99DE-1031636.
XX
XX 09-JUL-1999; 99DE-1032125.
XX
XX 09-JUL-1999; 99DE-1032126.
XX
XX 09-JUL-1999; 99DE-1032130.
XX
XX 09-JUL-1999; 99DE-1032186.
XX
XX 09-JUL-1999; 99DE-1032206.
XX
XX 09-JUL-1999; 99DE-1032227.
XX
XX 09-JUL-1999; 99DE-1032228.
XX
XX 09-JUL-1999; 99DE-1032229.
XX
XX 09-JUL-1999; 99DE-103230.
XX
XX 09-JUL-1999; 99DE-1032922.
XX
XX 14-JUL-1999; 99DE-1032926.
XX
XX 14-JUL-1999; 99DE-1032928.
XX
XX 14-JUL-1999; 99DE-1033004.
XX
XX 14-JUL-1999; 99DE-1033005.
XX
XX 14-JUL-1999; 99DE-1033006.
XX
XX 12-AUG-1999; 99US-0148613.
XX
XX 27-AUG-1999; 99DE-1040764.
XX
XX 27-AUG-1999; 99DE-1040765.
XX
XX 27-AUG-1999; 99DE-1040766.
XX
XX 31-AUG-1999; 99DE-1040832.
XX
XX 99DE-1041378.

```

PR   31-AUG-1999;      99DE-1041379.    99DE-1041380.
PR   31-AUG-1999;      99DE-1041389.    99DE-1041396.
PR   31-AUG-1999;      99DE-1041396.    99DE-1041396.
PR   03-SEP-1999;      99DE-1042076.    99DE-1042076.
PR   03-SEP-1999;      99DE-1042077.    99DE-1042077.
PR   03-SEP-1999;      99DE-1042086.    99DE-1042087.
PR   03-SEP-1999;      99DE-1042088.    99DE-1042088.
PR   03-SEP-1999;      99DE-1042095.    99DE-1042095.
PR   03-SEP-1999;      99DE-1042124.    99DE-1042124.
PR   03-SEP-1999;      99DE-1042129.    99DE-1042129.
xx   09-MAR-2000;      2000US-0187970.
xx                                     (BADI ) BASF AG.
PA
xx
PI   Pompejus M., Kroegeer B., Schroeder H., Zelder O., Haberhauser G.;
DR   WPI: 2001-137957/14.
DR   N-PSDB: AAF72186.
xx
PT   Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT   pathway proteins, useful for producing fine chemicals in
xx   microorganisms, including organic acids, nonproteinogenic amino acids,
xx   and purine and pyrimidine bases -
PS   Claim 20; Page 1354-1355; 1737bp; English.
xx
CC   AA771753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC   pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC   MP nucleic acids are useful for the production of fine chemicals
CC   in microorganisms, including organic acids, nonproteinogenic amino
CC   acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC   saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC   compounds, vitamins, cofactors, polyketides and enzymes.
xx
SQ   Sequence          312 AA:
Query Match                      19.5%; Score 337; DB 22; Length 312;
Best Local Similarity           30.3%; Pred. No. 1.6e-25;
Matches 96; Conservative 53; Mismatches 150; Indels 18; Gaps
OY   11 IGRSALLSRMLGTVNFSGRVDDDALRLMDHARDGINCLDTADMTGMRIKGTTEL 70
Db   2 YGSGLIYSRIVLIGSTVSGTSLAEAGDILFAINSAGTHLDVPNY---tlgvaem 57
OY   71 VGRWLAQGGGRREDIVTLKVGE----MSERVNSGSLARRHIASCEGLRLGVNDHD 126
Db   58 lgtmI-dvevrsavvijssegvpmlplgrivd---csrnlldqlvtlrlnlttyid 113
OY   127 VTQMHHIDRSAPFWDEVQAMDSLVAASKSVSYSSNFACMHIAA--AOENARRHSLGMY 184
Db   114 lwsyvgwdegprhphevadtlidyavrtrygyrvysqgwglavthaasnaasaarpv 173
OY   165 SHQCILYNLARRAELDELVLPAAQAYGLGFVPMSPERPLNGLLSGALEKTACTAVKSAGRQ 244
Db   174 vaqnaysllrraeagellptqhlvgftiaagrpilqgvyltakryseiphsraastgita 233
OY   245 VLIPPL---RPATAYEVEFCRNLEEDPAEVGLAWLSFGPGJAGAVTGPRPEOLDALK 300
Db   234 ewgsjldmgrlivaldaakgylgispaevtacttwrdtpgytavlvgartneshlik 293
OY   301 ASAMTIDEQAELSEDEI 317
Db   294 aesvtlpcrlgalddv 310

```

[illegible]

```

Db 177 aeyhmfgrekeveqplfhkigygamtwspiacqlvsg---kydsqipppyrasalkyq 233
OY 234 ---TWKSAQGAQVLLPSLRPAIEAYEKEFCRNLEGEPAEVLAWLSRPGIAGAVIGRR 290
Db 234 wlkdkllseegrtq-----gaklkelqalaeirlgctlpqlalawclrneqssvllgas 287
OY 291 TPEQDSALAKASAM--TLDEQALSELDER 317
Db 288 naeqimenaigaqlvlpklssisvheidst 316

```

RESULT 14 AAW69711

ID AAW69711 standard; Protein; 344 AA.

AC AAW69711;

DT 26-OCT-1998 (first entry)

DE Streptomyces clavuligerus protein sequence of orf1p1.

KW Streptomyces clavuligerus; bacterial gene; clavulanic acid; 5S clavum;
5R clavum; cas1; ORF; open reading frame; biosynthesis.

OS Streptomyces clavuligerus.

PN WO9833896-A2.

PD 06-AUG-1998.

PE 02-FEB-1998; 98WO-EP00644.

PR 04-FEB-1997; 97GB-0002218.

PA (SMIK) SMITHKLINE BEECHAM PLC.
(UYAL-) UNIV ALBERTA.

PI Anders C, Barton B, Griffin JP, Jensen S, Mosher RH;

DR WPI: 1998-437451/37.

DR N-PSDB: AAV50486.

PT DNA comprising defective 5S clavain biosynthesis gene(s) from
Streptomyces clavum - useful for producing clavulanic acid without
production of 5S clavum or clavam-2-carboxylates

PS Claim 3; Page 15-24; 29pp; English.

XX The present sequence represents the protein sequence of a gene specific
CC for 5S clavam biosynthesis from Streptomyces clavuligerus (SC) and which
CC is not essential for 5R clavam biosynthesis. The present invention also
CC describes: (1) a process for improving 5R clavam production in a
CC suitable microorganism comprising manipulation of DNA encoding as above
CC and its inclusion in the microorganism; (2) a process for improving 5R
CC clavam production in SC comprising disrupting or otherwise making
CC defective DNA regions flanking cas 1; (3) a process for the
CC identification of a microorganism suitable for high 5R clavam production
CC comprising a preliminary screening for microorganisms with low or no 5S
CC clavam production; (4) a microorganism which is capable of 5R clavam
CC production and low or no 5S clavam production obtainable by a process as
CC in (3); (5) clavulanic acid (CA) obtainable by the fermentation of a
CC microorganism as in (4), and (6) CA which is free of any 5S clavam; (7)
CC CA which is free of any clavam-2-carboxylate. The methods and products
CC can be used to produce organisms capable of producing increased amounts
CC of clavams suitably e.g. CA, for use as antibiotics. The methods can
CC also be used for the production of CA without the production of 5S
CC clavam or clavam-2-carboxylate.

XX Sequence 344 AA;

```

Query Match 19.0%; Score 329; DB 19; Length 344;
Best Local Similarity 33.4%; Pred. No. 1.1e-24;
Matches 105; Conservative 45; Mismatches 132; Indels 32; Gaps 8;

OY 23 LGTVNSGRVEDDDALRLMDHARDRGINCIDTADMGRLYKHNTELVGRMLAOGGRR 82
Db 38 lptcdtygrprataratatlaavdagvllldtcadqg-----lgagellgrava---grr 90
OY 83 EDTVLATKVGEMSERVNDGLSAR--HIIASCEGSLRRLGVHDIDVYOMHIDRSAPMD 140
Db 91 devliatkfgmrssdgaagqilcgepsyvracerslrrlgtdridllyghwtcpavple 150
OY 141 EYWMAMDLSVAGSKSYVSSNFAGCHHTAAQENARRRSLGMSQCIXNLAVRAELE 200
Db 151 etvgavaelvregekvyrrlglsps---aalrrdavahpvtav--qsewslwstgiede 204
OY 201 VLPAAQAYGLGVAFWSPHLGGLSGAL-----EKLAAGTAVKSAQ--GRAQVLLPQL 250
Db 205 vvpvcrelglgivaayaypigrflgtlrltdldgedfrrgqprfapalarnrslhlr 264
OY 251 RPAIEAYEKEFCRNLEGEPAEVLAWLSRPGIAGAVIGRTPEDLSALKASAMTLDEQA 310
Db 265 rpvad-----glglrlaqlawlhhrgeadvvpigtcampahladnlaasirldtrs 317
OY 311 LSELDEIFPPAVASG 324
Db 318 laevtaaispvsq 331

```

RESULT 15

ID AAB08255 standard; Protein; 344 AA.

AC AAB08255;

DT 04-DEC-2000 (first entry)

DE The 6C5 cell wall antigen of Candida albicans.

KW Cell wall protein; yeast infection; pathogenesis inhibitor; 6C5 antigen;
pathogenic yeast.

OS Candida albicans.

PN WO200048620-A1.

PD 24-AUG-2000.

PE 18-FEB-2000; 2000WO-US04228.

PR 19-FEB-1999; 99US-0120764.

PR 19-FEB-1999; 99US-0120765.

PA (UYVI-) UNIV VIRGINIA PATENT FOUND.

PI Hazen KC, Singleton DR, Masouka J, Wu JG, Glee PM;

DR WPI: 2000-565334/52.

DR N-PSDB: AAA63930.

PT New isolated hydrophobic protein antigens derived from the cell walls
of Candida albicans and their corresponding antibodies, useful for
detecting and treating yeast infection

PS Disclosure; Fig 5A-C; 62pp; English.

XX The present sequence represents a cell wall protein, designated 6C5, of
CC Candida albicans. The cell wall peptides and their corresponding
CC antibodies are useful for treating yeast infections in a patient,
CC as well as for preventing or detecting infection of any yeast
CC type. These include opportunistic and nosocomial infections. The
CC peptides are also useful in the development of peptidomimetics and/or
CC small organic molecules with therapeutic, diagnostic and/or research

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2001, 09:22:59 ; Search time 13.07 Seconds
(without alignments)
513.249 Million cell updates/sec

Title: US-09-463-705A-2
Perfect score: 1728
Sequence: 1 MTDPAAATHVRLGRSALTLSR.....IDETFPVAVSGGAPEAWLQ 333

Scoring table:
BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 193259 seqs, 20144635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgcn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgcn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgcn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgcn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgcn2_6/ptodata/2/1aa/PCRTUS_COMB.pep:*
6: /cgcn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1728	100.0	1114	2	US-08-576-626A-31 Sequence 31, Appl
2	356	20.6	329	2	US-08-606-143-1 Sequence 1, Appl
3	356	20.6	329	2	US-08-606-143-3 Sequence 3, Appl
4	331	19.2	329	2	US-08-606-143-2 Sequence 2, Appl
5	208.5	12.1	331	2	US-08-907-674-1 Sequence 1, Appl
6	208.5	12.1	331	2	US-09-215-087-1 Sequence 1, Appl
7	208.5	12.1	331	2	US-09-391-959-1 Sequence 1, Appl
8	196.5	11.4	327	3	US-08-907-674-3 Sequence 3, Appl
9	196.5	11.4	327	3	US-09-215-087-3 Sequence 3, Appl
10	196.5	11.4	327	3	US-09-391-959-3 Sequence 3, Appl
11	177	10.2	278	1	US-07-941-414-1 Sequence 1, Appl
12	177	10.2	278	1	US-08-249-377A-1 Sequence 1, Appl
13	177	10.2	278	1	US-08-585-595-1 Sequence 1, Appl
14	177	10.2	278	1	US-08-749-337-1 Sequence 1, Appl
15	161.5	9.3	318	2	US-08-336-198C-3 Sequence 3, Appl
16	157	9.1	277	1	US-08-585-595-2 Sequence 2, Appl
17	149	8.6	310	4	US-09-166-412-4 Sequence 4, Appl
18	140	8.1	309	4	US-09-166-412-2 Sequence 2, Appl
19	111.5	6.5	316	3	US-08-801-344-4 Sequence 4, Appl
20	108	6.2	323	3	US-08-853-839-2 Sequence 2, Appl
21	104	6.0	323	4	US-08-532-896-2 Sequence 2, Appl
22	99	5.7	316	1	US-08-585-595-3 Sequence 3, Appl
23	99	5.7	3170	2	US-07-642-734C-5 Sequence 5, Appl
24	97.5	5.6	3170	3	US-08-439-009A-5 Sequence 5, Appl
25	96.5	5.6	3567	2	US-09-036-987A-6 Sequence 6, Appl
26	96.5	5.6	3567	2	US-07-642-734C-4 Sequence 4, Appl
27	96.5	5.6	3567	3	US-08-439-009A-4 Sequence 4, Appl

28	95.5	5.5	7257	4	US-09-335-409-5	Sequence 5, Appl
29	93	5.4	431	1	US-08-391-939-18	Sequence 18, Appl
30	93	5.4	431	1	US-08-484-274A-18	Sequence 18, Appl
31	92.5	5.4	2257	1	US-08-611-107-10	Sequence 10, Appl
32	92.5	5.4	2257	2	US-08-422-560A-10	Sequence 10, Appl
33	92.5	5.4	2257	4	US-08-468-793-10	Sequence 10, Appl
34	92	5.3	1864	2	US-08-804-227C-3	Sequence 3, Appl
35	91.5	5.3	569	2	US-08-467-822-27	Sequence 27, Appl
36	91	5.3	4551	4	US-09-320-878-1	Sequence 1, Appl
37	89	5.2	407	2	US-08-926-327-2	Sequence 2, Appl
38	89	5.2	407	3	US-09-119-918-2	Sequence 2, Appl
39	88.5	5.1	402	2	US-08-403-852D-19	Sequence 19, Appl
40	88.5	5.1	402	3	US-08-510-646B-20	Sequence 20, Appl
41	88.5	5.1	402	4	US-09-231-818-19	Sequence 19, Appl
42	88.5	5.1	430	1	US-08-391-339-5	Sequence 5, Appl
43	88.5	5.1	430	1	US-08-484-274A-5	Sequence 5, Appl
44	87.5	5.1	1611	2	US-08-804-227C-5	Sequence 5, Appl
45	87.5	5.1	4472	2	US-08-804-227C-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-576-626A-31
Sequence 31, Application US/08576626A
Patent No. 5998194
GENERAL INFORMATION:
APPLICANT: Summers, R.G.
APPLICANT: Katz, L.
APPLICANT: Donadio, S.
APPLICANT: Staver, M.J.
TITLE OF INVENTION: POLYKETIDE-ASSOCIATED SUGAR
TITLE OF INVENTION: BIOSYNTHESIS GENES
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/576,626A
FILING DATE: 21-DEC-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dianne Casuco
REGISTRATION NUMBER: P-40,943
REFERENCE/DOCKET NUMBER: 5857.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (847) 938-3137
TELEFAX: (847) 938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1114 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5998194e
US-08-576-626A-31
Query Match 100.0%; Score 1728; DB 2; Length 1114;

Best Local Similarity 100.0%; Pred. No. 1.9e-178,
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MTTDAATVHRLGRSALLTSLRLMTGLTVNFSGRVEDDRLAMDHADRG INCLDPTADMGM	60
Db	782	MTTDAATVHRLGRSALLTSLRLMTGLTVNFSGRVEDDRLAMDHADRG INCLDPTADMGM	841
QY	61	RLTKHHTBELVGRMLAOGGGRREDVTVLTAKYGGENSERVNDSGLSARHTIASCESLRL	120
Db	842	RLTKHHTBELVGRMLAOGGGRREDVTVLTAKYGGENSERVNDSGLSARHTIASCESLRL	901
QY	121	GVDDHIDVYOMNHIDRSAPWDEWYQAMDLSLVASGKYSYGGSSNFAGMHTLAAQENARRHS	180
Db	902	GVDDHIDVYOMNHIDRSAPWDEWYQAMDLSLVASGKYSYGGSSNFAGMHTLAAQENARRHS	961
QY	181	LGMVSHOCLLYNLAVNAHAEVYLPAAQATGLGVFANSPLHGGILSALTEKLAAGTVKSAQ	240
Db	962	LGMVSHOCLLYNLAVNAHAEVYLPAAQATGLGVFANSPLHGGILSALTEKLAAGTVKSAQ	1021
QY	241	GRAQVLTSLRPALAEVEKEFCRNIGDEPAEYGLAVNLSRPAGIAGAVIPRTPEODLSALK	300
Db	1022	GRAQVLTSLRPALAEVEKEFCRNIGDEPAEYGLAVNLSRPAGIAGAVIPRTPEODLSALK	1081
QY	301	ASAMTIDEOALSDELDFPVAASGAAPAMIQ	333
Db	1082	ASAMTIDEOALSDELDFPVAASGAAPAMIQ	1114

RESULT 2

US-08-606-143-1
; Sequence 1, Application US/08606143
; Patent No. 5856155

```

1  GENERAL INFORMATION:
2  APPLICANT: Li, Min
3  TITLE OF INVENTION: COMPOUNDS AND RELATED METHODS FOR
4  TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH
5  TITLE OF INVENTION: COMPOUNDS
6  NUMBER OF SEQUENCES: 45
7  CORRESPONDENCE ADDRESS:
8  ADDRESSEE: Leydig, Volt & Mayer, Ltd.
9  STREET: Two Prudential Plaza, Suite 4900
10 CITY: Chicago
11 STATE: IL
12 COUNTRY: USA
13 ZIP: 60601
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: PatentIn Release #1.0, Version #1.25
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/606,143
21 FILING DATE: 23-FEB-1996
22 CLASSIFICATION: 514
23 ATTORNEY/AGENT INFORMATION:
24 NAME: Kilyk Jr., John
25 REGISTRATION NUMBER: 30763
26 REFERENCE/DOCKET NUMBER: 71756
27 TELECOMMUNICATION INFORMATION:
28 TELEPHONE: (312) 616-5600
29 TELEFAX: (312) 616-5700
30 TELEX: 25-3533
31 INFORMATION FOR SEQ ID NO: 1:
32 SEQUENCE CHARACTERISTICS:
33 LENGTH: 329 amino acids
34 TYPE: amino acid
35 STRANDEDNESS: single
36 TOPOLOGY: linear
37 MOLECULE TYPE: protein
38 US-08-606-143-1

```

Query Match	20.68; Score 356; DB 2; Length 329;
-------------	-------------------------------------

Best Local Similarity 30.98; Pred. No. 2.1e-30;
Matches 102; Conservative 57; Mismatches 131; Indels 40; Gaps 10;

Qy	11	IGRSLLTSTRMLCT--VNFSGRYVDDDLRLMHQABRG;NCLDTPDMYQWRLYKGTTEE	69
Db	4	LKRGSLVNSCGLCTWTFGGQIDVBAERKMTIATSSYGNLFDTAEVYA-----AGRAEV	59
Qy	70	LVGRMLAOGGCRREDYVLATKY--GGMSERYVNDGSLSANHTIASCBS;LRKGVCHIDY	127
Db	60	ILGSLIKKKGRRSSLVITTKLYGGKAE----ERGSLRHHIIIEGKLS;QOROLEVVDY	116
Qy	128	YQMHIDRSAPWDVQWQAMOSLVASGRYSYVSSGNFAGMHIIAAOENABRHSLGWYSRO	187
Db	117	VFAHRPDSNTPMEIIVRMTHIVINQGMAMTWGTSBRWAMEIMESYVARFNNIIPPCBQ	176
Qy	188	CLYNLAAYNH-AELEVLYAPAAQYIG;GVRAWSPBLHGGLLSGLALERTIACT-----	234
Db	177	AEYHFLFOREKEVEYVLP;ELYHKRIGAMTWSP;LAGIISG----KYGNQVPPSSRASIKCYO	233
Qy	235	----AVKSAQGRAOV-LPLSLRPAITAEYERKFCRNLGEDPAEVLGATWLSRPGIAGAYGP	289
Db	234	WLKERIVYSEBRKQONKIKD;SLPIAE-----RLGCTLPQLAIVAMWLNRGEGVSVLLGS	286
Qy	290	RTPEQDLSALKASAM--TLDEQALSEDEI	317
Db	287	STPQLLIENIGAIQVLEPRMTHSHVNEIDNI	316

RESULT 3

US-08-606-143-3
; Sequence 3, Application US/08606143
; Patent No. 5856155

1 GENERAL INFORMATION:
2 APPLICANT: Li, Min
3 TITLE OF INVENTION: COMPOUNDS AND RELATED METHODS FOR
4 TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH
5 TITLE OF INVENTION: COMPOUNDS
6 TITLE OF INVENTION: 45
7 NUMBER OF SEQUENCES: 45
8 CORRESPONDENCE ADDRESSES:
9 ADDRESSEE: Leydig, Volt & Mayer, Ltd.
10 STREET: Two Prudential Plaza, Suite 4900
11 CITY: Chicago
12 STATE: IL
13 COUNTRY: USA
14 ZIP: 60601
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: PatentIn Release #1.0, Version #1.25
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/606,143
22 FILING DATE: 23-FEB-1996
23 CLASSIFICATION: 514
24 ATTORNEY/AGENT INFORMATION:
25 NAME: Kilyk Jr., John
26 REGISTRATION NUMBER: 30763
27 REFERENCE/DOCKET NUMBER: 71756
28 TELECOMMUNICATION INFORMATION:
29 TELEPHONE: (312) 616-5600
30 TELEFAX: (312) 616-5700
31 TELEX: 25-3533
32 INFORMATION FOR SEQ ID NO: 3:
33 SEQUENCE CHARACTERISTICS:
34 LENGTH: 329 amino acids
35 TYPE: amino acid
36 STRANDEDNESS: single
37 TOPOLOGY: linear
38 MOLECULE TYPE: protein
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
77

Query Match	20.68; Score 356; DB 2; Length 329;
-------------	-------------------------------------

Best Local Similarity 30.9%; Pred. No. 2,1e-30;
Matches 102; Conservative 57; Mismatches 131; Indels 40; Gaps 10;

QY 11 LGRSALLTSRLMGT-VNFSGRVEDDALARMDHARDGINCIDTADMYGWRLYKGTTEE 69
DB 4 LKSGSLRVSCLGIGTWTGCGQISDEVALRMTIAGESVNFDAEYVA----AGKAV 59
QY 70 LVGRMLAOGGGRREDVTATKY--GGEMSERVNDGSLARHIIASCEGSLRLGVHDIV 127
DB 60 ILGSIKKKGRMRSSLVITTKLYWGKAE---TERLSKRHHIEGKSLQRLQLEVDV 116
QY 128 YOMHHIDRARPDEWQANDSLVASKSVYGSNPNAGHIIAAGQNAARRISLGWVSHQ 187
DB 117 VANRPDPTPMEIYRATHTVINOAMWTGTSRWSMEIWEAYSARQFNLIPPTCEQ 176
QY 188 CLYNLAVRH-AELEVLPAQAYGLGFVWSPHGLLSGALKELAAGT----- 234
DB 177 AEYHMFQRKKEVQVPELPHKIGVGAMTWSPLACGIVSG---KYSGIPYRSALIKYQ 233
QY 235 ---AVKSAQRAQV-LPSPRAIEAYEFCRNIGEDPAEVLAVYLSRPGIAGAVIGP 289
DB 234 WKERIVSEGRKQONKLDLSPAE-----RLGCTLPOLAVMCLNREGVSVLLGS 286
QY 290 RPEOLDALKASAM--TDEQALSELDEI 317
DB 287 STEPOLIENLGAIQVLPKMTSHVNEIDNI 316

RESULT 4

US-08-606-143-2
Sequence 2, Application US/08606143

Patent No. 5856155
GENERAL INFORMATION:
APPLICANT: LI, Min
TITLE OF INVENTION: COMPOUNDS AND RELATED METHODS FOR
TITLE OF INVENTION: MODULATING POTASSIUM ION CHANNELS AND ASSAYS FOR SUCH
TITLE OF INVENTION: COMPOUNDS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voit & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/606,143
FILING DATE: 23-FEB-1996
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Kilyk Jr., John
REGISTRATION NUMBER: 30763
REFERENCE/DOCKET NUMBER: 71756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
TELEX: 25-3533
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 329 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-606-143-2

Query Match

19.2%; Score 331; DB 2; Length 329;

Best Local Similarity 28.6%; Pred. No. 1,1e-27;
Matches 94; Conservative 60; Mismatches 137; Indels 38; Gaps 9;

QY 11 LGRSALLTSRLMGT-VNFSGRVEDDALARMDHARDGINCIDTADMYGWRLYKGTTEE 69
DB 4 LKSGSLRVSCLGIGTWTGCGQITDEMAMHMLATDNGINLFDIAEYVA----AKKAV 59
QY 70 LVGRMLAOGGGRREDVTATKY--GGEMSERVNDGSLARHIIASCEGSLRLGVHDIV 127
DB 60 VLGNIKKKGRMRSSLVITTKIFWGKAE---TERLSKRHHIEGKSLERLQLEVDV 116
QY 128 YOMHHIDRARPDEWQANDSLVASKSVYGSNPNAGHIIAAGQNAARRISLGWVSHQ 187
DB 117 VANRPDPTPMEIYRATHTVINOAMWTGTSRWSMEIWEAYSARQFNLIPPTCEQ 176
QY 188 CLYNLAVRH-AELEVLPAQAYGLGFVWSPHGLLSGALKELAAGT----- 233
DB 177 AEYHMFQRKKEVQVPELPHKIGVGAMTWSPLACGIVSG---KYSGIPYRSALIKYQ 233
QY 234 ---TAVKSAQRAQVLLPSPRAIEAYEFCRNIGEDPAEVLAVYLSRPGIAGAVIGP 290
DB 234 WKDKILSEGRRO-----QAKLKLQAIARLGCTLPOLAIAMCLNREGVSVLLGS 287
QY 291 RPEOLDALKASAM--TDEQALSELDEI 317
DB 288 NAEQLEMENIGAIQVLPKLSSTVHEIDSI 316

RESULT 5

US-08-907-674-1
Sequence 1, Application US/08907674

Patent No. 5919685
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/907,674
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0362 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:


```

REFERENCE/DOCKET NUMBER: PF-0362 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 331 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAINOT14
CLONE: 1596452
US-09-391-959-1

```

```

Query Match      12.1%; Score 208.5; DB 3; Length 331;
Best Local Similarity 23.9%; Pred. No. 2.2e-14;
Matches 83; Conservative 49; Mismatches 135; Indels 81; Gaps 10;

```

```

QY 23 LGTVNFGVEEDDLRLMDHARDGINCIDTADMTGWRLYKGTTELVGRWLAOGGRR 82
  11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 14 LGAMEGRMDAPTSAVTRAFLEGRHTEIDTAFVYS---EGQSEITIGLGLRLGGG 69
QY 83 EDTVLTAKYGGEMSERVNDGSLARHIIASCGSLRLGLVDHIDVYQMHIDRSAPWDE 142
  11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 70 CRYATKTK-----ANPMDKSLKPDSEVRSQLETSKRLQCPQVDLYLTPHGTVEET 124
QY 143 WQAMDSLVAAGKSYVSSNFGMHIIAAOENARHSHSGMV---SHOCLYNLAVRHAEL 199
DB 125 LIAQGLHDEGKFEVGLSLNYSWEVA---EICTLCKSNQWIIPTYYQGMNATITQVET 181
QY 200 EYLPAQAAYGLGVFANSPHGLLSGAL-----EKL 230
  11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 182 ELFPCLRHGFLRFYAVNPLAGLLTGKYEKDKQPVGRFGNTWAEMRYRMKKEHH 241
QY 231 AAGTAV--KSAQRAQVLLPSLRPALEAEKFCRNIGEDPAEGLAWLSRPIAQA--- 285
DB 242 FESIALVEALQAAYGASAPSVTSA-----ALWMMHHSQLOGAHD 283
QY 286 --VIGRPEQLDSALKASAMTLEQATSELEIFPAPVAVASGAAPPAW 331
DB 284 AVILGMSLEQLLEONL-----AATEEGLEPAVVD--AFNQAW 319

```

```

RESULT 8
US-08-907-674-3
Sequence 3, Application US/08907674
Patent No. 5919685
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN AFLATOXIN BI ALDEHYDE REDUCTASE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/907,674
FILING DATE: Herewith
CLASSIFICATION: 514

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0362 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 433611
US-08-907-674-3

```

```

Query Match      11.4%; Score 196.5; DB 2; Length 327;
Best Local Similarity 23.5%; Pred. No. 4.2e-13;
Matches 80; Conservative 49; Mismatches 146; Indels 65; Gaps 12;

```

```

QY 23 LGTVNFGVEEDDLRLMDHARDGINCIDTADMTGWRLYKGTTELVGRWLAOGGRR 82
  11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 10 LGAMEGRMDAPTSSASVAFIQRGHTIEDTAFVYA---NQSEITIGD-LGLIGRS 64
QY 83 EDTV-LATKVGEMSERVNDGSLARHIIASCGSLRLGLVDHIDVYQMHIDRSAPWDE 141
  11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 65 GCRVKATRAAPMGKTK-----LKPADVRFQLETSKRLQCPQVDLYLTPHGTVEET 119
QY 142 WQAMDSLVAAGKSYVSSNFGMHIIAAOENARHSHSGMVSHOCLYNLAVRHAEL 201
DB 120 TLQACHVHDEGKFEVGLSLNYSWEVAELCTLCKNNGMIMPVYQGMNATITQVETEL 179
QY 202 LPAQAAYGLGVFANSPHGLLSGALAKLAAGTAVKSAQRAQVLLPSLR---PAIEAV 257
  11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 180 FPLCLRHGFLRFYAVNPLAGLLTGRYK-----YQDKQDN---PESRFGNPSQLY 228
QY 258 -----EKFCRNIGEDPAEGLAWLSRPIAQAIVIGRPEQLDSALK----- 300
  11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
DB 229 MDRYKKEHF-----NGIALV---EKALKTTYGTPAPSMISAAYRMVYHNSQLNG 275
QY 301 --ASAMTLEQALSELDEIFPAPVAVASGAAP-----EAW 331
DB 276 TQGDVAIIGMSLEQLLEONLALVEEGPLEPAPVAVDARDQAW 315

```

```

RESULT 9
US-09-215-087-3
Sequence 3, Application US/09215087
Patent No. 5981244
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN AFLATOXIN BI ALDEHYDE REDUCTASE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible

```

```
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/215,087
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/907,674
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0362 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 433611
US-09-215-087-3
```

```
Query Match 11.48; Score 196.5; DB 2; Length 327;
Best Local Similarity 23.58; Pred. No. 4.2e-13;
Matches 80; Conservative 49; Mismatches 146; Indels 65; Gaps 12;
```

```
QY 23 LGTVNFSGRVEDDRLIMHARDGINCLDTADMYGMRLYKGTETELVGMWLAOGGRR 82
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 10 LGAMEMGRNDVYSSASVRFQRGHTEIDTAHYA-----NGOSEITLGD-LGLGLGRS 64
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 83 EDVY-LATKYGGENSERVNDGSLARHIIASCEGSLRLRGVDHIDVYOMHHIDRSAPWDE 141
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 65 GCKVKIATKAPMGKGT-----LKPADVRFQLETSLKRLCCPRVDLYLHPDHGTFIEE 119
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 142 VMQAMDSLVASGKSYVSSNFAGMHIIAAQENARHSLGMSHOCILYNLAHNALEEV 201
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 120 TLQCHHVHOGKFEVGLSLNYSWEVAEICTLCKKNGWIMPTVYQGMYNATIRQVETEL 179
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 202 LPAQAYGLGVFAMSPPLHGILSGALEKLAAGTAVKSAOGRAGVLLPSLR-----PAIEAY 257
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 180 FPCLRHFGRLRYAFNPLAGILGILGRK-----YQKDKGN-----PESRFGGPFQOLY 228
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 258 -----EKFCRNIGEDPAEYGLAWLSRPGIAGAVIGPTPEQLDSALK----- 300
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 229 MDRYWKKEHF-----NGIALV---EKALKTYGPTAPSMISAAYVWMYHHSQLKG 275
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 301 --ASAMTLDEQALSELDEIFPAVASGGAAP-----EAW 331
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 276 TQGDVAILGMSLSLEQLEONLALVEEGPLEPVAVDAPDOAW 315
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
```

```
RESULT 10
US-09-391-959-3
Sequence 3, Application US/09391959
Patent No. 6071704
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Shah, Puryi
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
```

```
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/391,959
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/907,674
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0362 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 433611
US-09-391-959-3
```

```
Query Match 11.48; Score 196.5; DB 3; Length 327;
Best Local Similarity 23.58; Pred. No. 4.2e-13;
Matches 80; Conservative 49; Mismatches 146; Indels 65; Gaps 12;
```

```
QY 23 LGTVNFSGRVEDDRLIMHARDGINCLDTADMYGMRLYKGTETELVGMWLAOGGRR 82
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 10 LGAMEMGRNDVYSSASVRFQRGHTEIDTAHYA-----NGOSEITLGD-LGLGLGRS 64
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 83 EDVY-LATKYGGENSERVNDGSLARHIIASCEGSLRLRGVDHIDVYOMHHIDRSAPWDE 141
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 65 GCKVKIATKAPMGKGT-----LKPADVRFQLETSLKRLCCPRVDLYLHPDHGTFIEE 119
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 142 VMQAMDSLVASGKSYVSSNFAGMHIIAAQENARHSLGMSHOCILYNLAHNALEEV 201
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 120 TLQCHHVHOGKFEVGLSLNYSWEVAEICTLCKKNGWIMPTVYQGMYNATIRQVETEL 179
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 202 LPAQAYGLGVFAMSPPLHGILSGALEKLAAGTAVKSAOGRAGVLLPSLR-----PAIEAY 257
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 180 FPCLRHFGRLRYAFNPLAGILGILGRK-----YQKDKGN-----PESRFGGPFQOLY 228
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 258 -----EKFCRNIGEDPAEYGLAWLSRPGIAGAVIGPTPEQLDSALK----- 300
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 229 MDRYWKKEHF-----NGIALV---EKALKTYGPTAPSMISAAYVWMYHHSQLKG 275
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
QY 301 --ASAMTLDEQALSELDEIFPAVASGGAAP-----EAW 331
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
DB 276 TQGDVAILGMSLSLEQLEONLALVEEGPLEPVAVDAPDOAW 315
||| :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|: :|:
```

```
RESULT 11
US-07-941-414-1
Sequence 1, Application US/07941414
Patent No. 5376544
GENERAL INFORMATION:
APPLICANT: LAZARUS, ROBERT A.
APPLICANT: HURLE, MARK
APPLICANT: ANDERSON, STEPHEN
APPLICANT: POWERS, DAVID B.
```

RESULT: 12
US-08-249-377A-1
Sequence 1, Application US/08249377A
: Patent No. 5563025
: GENERAL INFORMATION:
: APPLICANT: LAZARUS, ROBERT A.
: APPLICANT: HURLE, MARK
: APPLICANT: ANDERSON, STEPHEN
: APPLICANT: POWERS, DAVID B.

RESULT 13
US-08-585-595-1
Sequence 1, Application US/08585595
Patent No. 5795761
GENERAL INFORMATION:
APPLICANT: POWERS, DAVID B.
APPLICANT: ANDERSON, STEPHEN
TITLE OF INVENTION: IMPROVED METHODS FOR PRODUCING VITAMIN C
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:

ADDRESSEE: HOWREY & SIMON
STREET: 1299 PENNSYLVANIA AVENUE, N.W.
CITY: WASHINGTON
STATE: DC
COUNTRY: US
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/585,595
FILING DATE: 16-JAN-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/584,019
FILING DATE: 11-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: AUERBACH, JEFFREY I.
REGISTRATION NUMBER: 32680
REFERENCE/DOCKET NUMBER: 6137-0014 CIP
TELEPHONE: (202) 383-7451
TELEFAX: (202) 383-6610
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: 2,5 DKG REDUCTASE A
INDIVIDUAL ISOLATE: CORYNEBACTERIUM SP.
US-08-585-595-1

Query Match 10.2%; Score 177; DB 1; Length 278;
Best Local Similarity 25.1%; Pred. No. 4.3e-11;
Matches 76; Conservative 44; Mismatches 117; Indels 66; Gaps 11;
QY 31 RVEDDDALRLMDHARDGNCINCLTDADMYGRLYKGTIELVYGRRLAOGGREGRETVLATK 90
DB 23 KVPADPQRAVEALEVEGYRHIDTAIYG-----NEEGGAIAASGIARDLDFITTK 75
QY 91 VGGEMSERVNDGSLASR-----IIASCEGSLRGLGVHDIVYOMHHIDRSAP-WDEVWQA 145
DB 76 LM-----ND-----RHGDDEPAATAESLAKLADQVDLYVHPTPPAADNVYHAMEK 123
QY 146 MDSLVAAGKVSYYGSSNFAGMHTAAQENAAARRHSLGMSHOCLYNLAVRHAEEVLPA 205
DB 124 MIELRAAGLRLRSIGSVN-----HLVPHLERIVA--ATGVVPAVNOIELHPAYQOREITDMA 177
QY 206 QAVGLGFANSPHLGCL--LSGALEKLAAGTAVKSAOGRQVLLPSLRPAIEAYEYFCRN 263
DB 178 AAHDVKRIESGPIGQGYKDYLFGAEPVTAATAA-----209
QY 264 LGEDPAEVLGLAWLVRGICGAVIGPRT--PEQDLSALKASAMTLDQALSELDEIFPAV 321
DB 210 HGKTPAQAVLRMHLQK---GFVVEPKSVRRERLEENLDVDFEDLTDTETIAAIDAMPDGD 265
QY 322 ASG 324
DB 266 GSG 268

RESULT 14
US-08-749-337-1
Sequence 1, Application US/08749337
Patent No. 5912161
GENERAL INFORMATION:
APPLICANT: LAZARUS, ROBERT A.

APPLICANT: HURLE, MARK
APPLICANT: ANDERSON, STEPHEN
APPLICANT: POWERS, DAVID B.
TITLE OF INVENTION: ENZYMES FOR THE PRODUCTION OF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: HOWREY & SIMON
STREET: 1299 PENNSYLVANIA AVE., N.W.
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/749,337
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: AUERBACH, JEFFREY I.
REGISTRATION NUMBER: 32,680
TELEPHONE: (202) 383-7451
TELEFAX: (202) 383-6610
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 278 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-749-337-1

Query Match 10.2%; Score 177; DB 2; Length 278;
Best Local Similarity 25.1%; Pred. No. 4.3e-11;
Matches 76; Conservative 44; Mismatches 117; Indels 66; Gaps 11;
QY 31 RVEDDDALRLMDHARDGNCINCLTDADMYGRLYKGTIELVYGRRLAOGGREGRETVLATK 90
DB 23 KVPADPQRAVEALEVEGYRHIDTAIYG-----NEEGGAIAASGIARDLDFITTK 75
QY 91 VGGEMSERVNDGSLASR-----IIASCEGSLRGLGVHDIVYOMHHIDRSAP-WDEVWQA 145
DB 76 LM-----ND-----RHGDDEPAATAESLAKLADQVDLYVHPTPPAADNVYHAMEK 123
QY 146 MDSLVAAGKVSYYGSSNFAGMHTAAQENAAARRHSLGMSHOCLYNLAVRHAEEVLPA 205
DB 124 MIELRAAGLRLRSIGSVN-----HLVPHLERIVA--ATGVVPAVNOIELHPAYQOREITDMA 177
QY 206 QAVGLGFANSPHLGCL--LSGALEKLAAGTAVKSAOGRQVLLPSLRPAIEAYEYFCRN 263
DB 178 AAHDVKRIESGPIGQGYKDYLFGAEPVTAATAA-----209
QY 264 LGEDPAEVLGLAWLVRGICGAVIGPRT--PEQDLSALKASAMTLDQALSELDEIFPAV 321
DB 210 HGKTPAQAVLRMHLQK---GFVVEPKSVRRERLEENLDVDFEDLTDTETIAAIDAMPDGD 265
QY 322 ASG 324
DB 266 GSG 268

RESULT 15
US-08-336-198C-3
Sequence 3, Application US/08336198C
Patent No. 586382
GENERAL INFORMATION:
APPLICANT: Hallborn, Johan
APPLICANT: Penttila, Merja

APPLICANT: Ojamo, Heikki
APPLICANT: Keranen, Sirkka
APPLICANT: Hahn-Hagerdal, Barbel
APPLICANT: Waldfidsson, Mats
APPLICANT: Airaksinen, Ulla
TITLE OF INVENTION: Xylose utilization by recombinant yeasts
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 N. Washington St.
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22046-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,198C
FILING DATE: 03-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
TELEX: 248345
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 318 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-198C-3

Query Match 9.3%; Score 161.5; DB 2; Length 318;
Best Local Similarity 23.5%; Pred. No. 2.5e-09;
Matches 70; Conservative 49; Mismatches 104; Indels 75; Gaps 12;
QY 59 GRRLYKG---HTELVG---RWLAOGGGRREDYVLTATKVGEMSERVNDGSLSRHIT 110
DB 38 GYRLFDGAEVDYANEKLVAGVKKAIDEGIVKREDLFLTSKLMNNYHHPDN-----VE 89
QY 111 ASCEGSLRLGVHDHIDVOMH-----HID-RSAPNDEVNQA 145
DB 90 KALNRTLSDLDVDYDLFLHPVTFKFPVLEEKYPPGFGCGKDNFDYEDVPILETWKA 149
QY 146 MDSLVASKVSVYSSNPFAGMIIAAQENAAARHSISGMVSHOCLYNLAVRHALE---VL 202
DB 150 LEKLVKAKIRISIGVSNPGALLDLRLGATIKPSVLQVEH-----HPYLQOPRLI 200
QY 203 PAQAYIGLCVRFMSPLHGLLSGALEKLAGTAVKSAQGRAOVLLPSLRPAIEAYEKFCR 262
DB 201 EFAOSRGIAVTAYS-----SFGPOSFVELNQGRLNLTSPLEFN--ETIKATAA 246
QY 263 NIGEDPAVEGLAWLVSFPGIAGAVTGPR---PEQLDSALKASAMTLDQALSELDEI 317
DB 247 KHGKSPAOVLLRWSSOR---GIALIPKSNVYPRILENK-DVNSFPLDQDFADIATKL 299

Search completed: June 22, 2001, 09:24:49
Job time: 110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 23, 2001, 10:02:37 ; Search time 1502.48 Seconds

(without alignments)
10284.517 Million cell updates/sec

Title: US-09-463-705A-1_COPY_48_1046

Perfect score: 999

Sequence: 1 ctcgaaccgcgctccgcgc9.....gtccgcgcgtcgtgtgtcat 999

Scoring table: IDENTITY_NUC

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgc_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_hum8:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*
96: gb_in4:*
97: gb_pr10:*
98: em_ba3:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length-DB	ID	Description
1	999	100.0	3439	SEV14332
2	999	100.0	3439	AX000482
3	999	100.0	3480	U77434 Saccharopol
4	999	100.0	3756	AR092721
5	673.4	67.4	47981	AF263245
6	593.6	59.4	5908	AF147704
7	511.6	51.2	9530	AB032523
8	486.2	48.7	47981	AF263245

Description

Y14332 Saccharopol
AX000482 Sequence
U77434 Saccharopol
AR092721 Sequence
AF263245 Micromono
AF147704 Streptomy
AB032523 Streptomy
AF263245 Micromono

c	9	374	37.4	22991	3	SCD19	AL392149	Streptomy
	10	346.4	34.7	30657	3	SCD12A	AL357524	Streptomy
	11	314.8	31.5	12335	1	AE003996	AE003996	Xylella f
	12	213.2	21.3	34074	3	SCF81	AL133171	Streptomy
	13	208.4	20.9	39896	3	SCK13	AL451182	Streptomy
c	14	185.2	18.5	10314	1	AE004600	AE004600	Pseudomon
c	15	180.8	18.1	349116	2	AP003003	AP003003	Mesorhizo
c	16	179	17.9	8058	3	SCB56	AL138852	Streptomy
c	17	167.4	16.8	18495	2	BS093875	U93875	Bacillus su
c	18	167.4	16.8	213420	2	BS090014	299117	Bacillus su
c	19	167	16.7	145914	2	AP003014	AP003014	Mesorhizo
c	20	157.6	15.8	22134	3	SC152	AL590507	Streptomy
c	21	150	15.0	1512	10	AX078635	AX078635	Sequence
c	22	147.2	14.7	12257	1	AE002963	AE002063	Deinococc
c	23	140.4	14.1	17148	1	AE004798	AE004798	Pseudomon
c	24	139.6	14.0	36583	3	SCSH1	AL049863	Streptomy
c	25	139.4	14.0	349498	2	AP003002	AP003002	Mesorhizo
c	26	135.6	13.6	7449	3	SCD17A	AL392177	Streptomy
c	27	132.4	13.3	13586	3	SC7H9	AL450223	Streptomy
c	28	132.4	13.3	194140	2	AF242881	AF242881	Agrobacte
c	29	129.6	13.0	10016	1	AE005035	AE005035	Halobacte
c	30	129	12.9	12021	1	AE005073	AE005073	Halobacte
c	31	117	11.7	13291	1	AE005932	AE005932	Calubact
c	32	113.6	11.4	39896	3	SCK13	AL451182	Streptomy
c	33	113.4	11.4	10948	1	AE004857	AE004857	Pseudomon
c	34	112.4	11.3	5413	3	HV095374	U95374	Haloferax v
c	35	111.8	11.2	346897	2	AP002995	AP002995	Mesorhizo
c	36	108.2	10.8	984	14	EDKCBETA	AL225806	Egeria de
c	37	108.2	10.8	11104	1	AE005963	AE005963	Calubact
c	38	104.4	10.5	10307	1	AE004542	AE004542	Pseudomon
c	39	96.6	9.7	345783	2	AP003001	AP003001	Mesorhizo
c	40	95.4	9.5	13184	1	AE005096	AE005096	Halobacte
c	41	95.2	9.5	332635	2	AP003005	AP003005	Mesorhizo
c	42	92.4	9.2	347750	2	AP002998	AP002998	Mesorhizo
c	43	92	9.2	1032	9	A89638	A89638	Sequence 4
c	44	92	9.2	2929	9	SCCLAVAM1	AF124928	Streptomy
c	45	92	9.2	7193	9	A89635	A89635	Sequence 1

ALIGNMENTS

RESULT	1	SEV14332	3439 bp	DNA	BCT	24-JUL-1998
LOCUS		SEV14332				
DEFINITION		Saccharopolyspora erythraea partial eryG and eryA genes and eryBII, eryCII and eryCII genes.				
ACCESSION		Y14332.1	GI:2765405			
VERSION		eryA gene; eryBII gene; eryCII gene; eryCIII gene; eryC gene.				
KEYWORDS		Saccharopolyspora erythraea.				
SOURCE		Saccharopolyspora erythraea				
ORGANISM		Bacteria: Firmicutes: Actinobacteria: Actinobacteridae: Actinomycetales: Pseudonocardiales: Pseudonocardaceae; Saccharopolyspora				
REFERENCE		1 (bases 1 to 3439)				
AUTHORS		Salah-Bey, K., Douthett, M., Michel, J.M., Haydock, S., Cortes, J., Leadlay, P.F. and Raynal, M.C.				
TITLE		Targetted gene inactivation for the elucidation of deoxysugar biosynthesis in the erythromycin producer saccharopolyspora erythraea				
JOURNAL		Unpublished				
REFERENCE		2 (bases 1 to 3439)				
AUTHORS		Leadlay, P.F.				
TITLE		Direct Submission				
JOURNAL		Submitted (28-JUL-1997) P.F. Leadlay, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QM, UK				
FEATURES		Location/Qualifiers				
source		1..3439				
gene		/organism="Saccharopolyspora erythraea"				
		/strain="NRRL2338"				
		/db_xref="taxon:1836"				
		/complement(1..13)				

CDS	/gene="eryG"	/complement(<1..13)	/gene="eryG"	/function="Involved in erythromycin A biosynthesis"
	/codon_start=1			
	/transl_table=11			
	/product="O-methyltransferase"			
	/protein_id="CAA74708.1"			
	/db_xref="GI:4388593"			
gene	/translation="MSVK"			
	/complement(45..1046)			
	/gene="eryBII"			
CDS	/complement(45..1046)			
	/function="Involved in mycarose biosynthesis"			
	/note="putative"			
	/codon_start=1			
	/transl_table=11			
	/product="dtp-4-keto-L-6-deoxy-hexose 2,3-reductase"			
	/protein_id="CAA74709.1"			
	/db_xref="GI:2765406"			
	/db_xref="SPTREMBL:O33936"			
	/translation="MTDAATHTVGRSALLTSRLTGTVNFSGVEPDADLRMDHA			
	RDRGINDPTADMYGRRLKGTBELVGMHLAOGGREDPTVLATKVGENSERNDG			
	GLSARHIIASCEGSLRGLVDHIDYQMHIDRSAPWEVWGMASLAVSGKVSYS			
	SNFAGMHIAAOENARHSLGMSHOCLYNLAHVHAELEVLPAQAQGLAVFAMSP			
	HGGLSLGALKELACTAVKSAQRAQVLLPSLRPAIEAEKFCRMIGEDPAVGLAW			
	LSRPGIAGVIVCPTRPEQDLSAKMSAMTLDQALSELDEIRPVAASGGAPEAMVLO"			
gene	/complement(1043..2308)			
	/gene="eryCIII"			
	/complement(1043..2308)			
CDS	/gene="eryCIII"			
	/function="Involved in desosamine biosynthesis"			
	/note="putative"			
	/codon_start=1			
	/transl_table=11			
	/product="glycosyltransferase"			
	/protein_id="CAA74710.1"			
	/db_xref="GI:2765407"			
	/db_xref="SPTREMBL:O54224"			
	/translation="MRVFSMSASKSLFGLVPLAMAFRAGCHEVAVVAPATFEDIT			
	ANGIRAVPGTVDLDVDFTHAGHDIIIDVRSIDSESDPATLTMEHLMOGTLEPT			
	FYALSPDTLIGMVSFCRRKRPDLTWEPLTFAPAIATVGTTHARLMQDPTTR			
	ARQNFLLGLDLPDEHREDPLAEMLTWLEKYGAPFEEVVGQWTLDPAPAAALR			
	TDKLTGVMRVYVNGSVYPEVILHDEPERRVCILTIGISSRENSIGOVSIIEILGAV			
	GDVAETIAFDAAOLEGVANIPDNRTVGVFMHALLPCATVTHHGGSGSWHTAIIH			
	GVPOVILPDGMPTGVRAORTQFEGGIALPVELRPDQRESVKRVLDPPAIRAGAR			
	KMDDMAERSPAEVVIGICELAAGRERR"			
gene	/complement(2319..3404)			
	/gene="eryCII"			
	/complement(2319..3404)			
CDS	/gene="eryCII"			
	/function="Involved in desosamine biosynthesis"			
	/note="putative"			
	/codon_start=1			
	/transl_table=11			
	/product="dtp-4-keto-6-deoxy-hexose 3,4-isomerase"			
	/protein_id="CAA74711.1"			
	/db_xref="GI:2765408"			
	/db_xref="SPTREMBL:O54225"			
	/translation="MTTDRAGLGRLOMIRCLHMGYSCNCPYMLCGHDDPQGR			
	TRSMRSGVRSRTLETWVADHATAROVLDPAFTATGRTPEMRRAAGAPPAEAOP			
	FRDVHAASWEGSPVGLAESFAGILPGAGRLDVGDFAVQVVGQWTLAVLGAAGV			
	LRGAAMDARVSLDAOLSPQOLAVTEAAVAALPADPALALFAGAEHTANTVDAVLAV			
	SAEPGLAERIADPPAAOATVAEVLRLHPLALHLEERTATAEVRLGEHYGEEEVVVV			
	VVAANRDEPVEFAPDRDIDRDPADRALSAHGHGRLEELVLTALATALRAANAALP			
	GLTPSGPVYRRRSRYLGRNCPVEL"			
gene	/complement(3401..3439)			
	/gene="eryA"			
	/complement(3401..3439)			
CDS	/gene="eryA"			
	/function="Involved in erythromycin A biosynthesis"			
	/codon_start=1			


```

|||||
108 TTGCGACAGCGCTGCTGTCAGAGGTCATCGCGAGCGCTTCAAGCCGAGTCGACCTG 167
121 CTGCGGAGTTCGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCG 180
168 CTGCGGAGTTCGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCG 227
181 GAGCGCGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 240
228 GAGCGCGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 287
241 GCGCGGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 300
288 GCGCGGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 347
301 GCGCGGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 360
348 GCGCGGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 407
361 GAGGAGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 420
408 GAGGAGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 467
421 GCGCGGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 480
468 GCGCGGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 527
481 CTCTGCGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 540
528 CTCTGCGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 587
541 GCGCGGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 600
588 GCGCGGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 647
601 GATGAGTGTGATGCTGAGAGCGTGTGATGAGAGCGTGTGATGAGAGCGTGTGATGAGAG 660
648 GATGAGTGTGATGCTGAGAGCGTGTGATGAGAGCGTGTGATGAGAGCGTGTGATGAGAG 707
661 GCGAGAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
708 GCGAGAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 767
721 GCGCGGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGG 780
768 GCGCGGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGG 827
781 CTCTGCGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 840
828 CTCTGCGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 887
841 GCGCGGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGG 900
888 GCGCGGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGG 947
901 CTCTGCGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGG 960
948 CTCTGCGCGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGGCGGCGG 1007
961 GCGCGGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGG 999
1008 GCGCGGCGGCGGCGGATGACGCGCGATCGCGGCGCGGCGCGGCGCGGCGGCGGCGG 1046

```

```

RESULT 3
LOCUS SEU77454 3480 bp DNA BCT 10-NOV-1997
DEFINITION Saccharopolyspora erythraea DEBSIII (eryCIII) and erythromycin
O-methyltransferase (eryC) genes, partial cds, and ERYCII (eryCII),
desosaminyltransferase (eryCII) and ERYBII (eryBII) genes, complete
cds.
ACCESSION U77454
VERSION U77454.1 GI:2599274

```

```

KEYWORDS
SOURCE Saccharopolyspora erythraea.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Pseudonocardiales; Pseudonocardaceae;
Saccharopolyspora.
REFERENCE
1 (bases 1 to 3480)
Summers, R.G., Donadio, S., Staver, M.J., Wendt, Pienkowski, E.,
Hutchinson, C.R. and Katz, L.
Sequencing and mutagenesis of genes from the erythromycin
biosynthetic gene cluster of Saccharopolyspora erythraea that are
involved in L-mycarose and D-desosamine production
Microbiology 143 (Pt 10), 3251-3262 (1997)
2 (bases 1 to 3480)
Summers, R.G., Staver, M.J., Donadio, S., Wendt, Pienkowski, E.,
Hutchinson, C.R. and Katz, L.
Direct Submission
Submitted (05-NOV-1996) D-47N AP9A, Abbott Laboratories, 100 Abbott
Park Road, Abbott Park, IL 60064, USA
FEATURES
source
/organism="Saccharopolyspora erythraea"
/strain="NRRL 2338"
/db_xref="taxon:1836"
1. 57
/gene="eryAIII"
<1. 57
/gene="eryAIII"
/codon_start=1
/transl_table=11
/product="DEBSIII"
/protein_id="AB84065.1"
/db_xref="GI:2599275"
/translation="HADAIARHIDAWLGGSNS"
43. 1136
/gene="eryCII"
43. 47
/gene="eryCII"
/feature="putative"
54. 1136
/gene="eryCII"
/codon_start=1
/transl_table=11
/product="ERYCII"
/protein_id="AB84066.1"
/db_xref="GI:2599276"
/translation="MTTTPDRAGCIGROLQMIKGLHMGYSNGDPYMLLGHDDPPOR
YKMSRSGVRRRTETWVADHATROVLDDPATRTATGRTPEMRRAGAPPEWAOPE
RDVHAASWGEVDPVDELAESEFGLPGAGARLDVGDGPVQGMVAIVGAAGVL
RGAAWADARVSLDQSPQOLAVTEAVALPAPDALPAGAEKMTANTVAVLAVS
AEPGLAEIADDPAAOARTVAEVLRLPALHERRATAEVRLGEHVIIEGEEVVVVV
AAANRDEVEAEPRDRIDVDRPADRALSHRHPRLBELVYALALRAAKALPG
LTPSGPVVRRRSPVLRGTRCVEL"
1135. 2412
/gene="eryCIII"
1135. 1138
/gene="eryCIII"
/feature="putative"
1147. 2412
/gene="eryCIII"
/codon_start=1
/transl_table=11
/product="desosaminyltransferase"
/protein_id="AB84067.1"
/db_xref="GI:2599277"
/translation="MRVYSSMSKSHLFLGLPPLAFAAGAEHVAVASPALEDTIT
AAGLTVAPGTVDDVDFMTWACHDIIIDVRSIDFSEKRPATLTWHLGMQVLPPT
FVFLMSPDRLIEGMSFCKRMKRPDLIWEPLTFEAPAIAAVGTPHARLGMGPDITTR
ARONFLGLPDQDEEHEGRLAEMLTWLEKYGGAPEDEEVVVGWITIDPAPAIRD
TGLKTVGMRYVDNPSVVPENMLHDEPRRRVCTLTGISRENSICVSIIEELGAVG
DVDAEIIATFPDAQLEGSVANIPIHNVATGVFVRPHALLPTPCARIVYHHGSGSWHTAIH
GPOVILIPGMDTGVRARQKROERFAGIALFVPELTPDOLRESVKRYLVHGDPAHRAGAR

```

```

gene      2394..3410
/gene="eryB1"
RBS      2394..2399
/gene="eryB1"
CDS      2409..3410
/note="putative"
/gene="eryB1"
/codon_start=1
/transl_table=11
/product="EryB1"
/protein_id="AAB84068.1"
/db_xref="GI:2599278"
/translation="MTTDAATHTVRLGSSALLTSRLWLTGVNFSGRVEDDADLRIMDADR
KDRINCINCLDTADMTGWRLYKGTETELVGRMTAOGGRRREDIVLTKVGGSESEVNSD
GLSRHIIASCEGSLRLRGYADIVYDMHIIDRSAPDEWQADSLVASGKSVYSS
SNFGNIAAQAENARRSHSLGANSHOCLYLAVRAIELEVLPAQAQYGLGFVAFSPIL
LSRGISGLEKLEAGTAGTVKSAQGRAOYILPILRAPIEAIEYERCFNIGEDDPAEYGLAWY
LSRGISGAVTIGPRTPEQDLSALKAASMTLDEQALSELDELTFPAVASGARPEMWLQ"
4429..3480
/gene="eryG"
RBS      3429..3433
/gene="eryG"
CDS      3442..>3480
/note="putative"
/gene="eryG"
/codon_start=1
/transl_table=11
/product="erythromycin O-methyltransferase"
/protein_id="AAB84069.1"
/db_xref="GI:2599279"
/translation="MSVKQKSAIOLDLV"

```

[illegible]

Db	2927	CTCCTGGCGGGCGGCGATGTGCCAGACCCGGGAAGTTGACGAGCCGACCTAGACACTT	2868
QY	541	gcgcctgacgacgagagctgtcatgtgccttccacactcgtcccaacgycgcgacgagtc	600
Db	2867	GCCGCTGGCGACGAGGGCTGTTCATAGGCTCTCCACACCTGTGTCACAGGGGCGACCGCT	2808
QY	601	gattgtgtgtcatctgtgtagacgttcgaatgtgtgttcgaccccaagctcgcgcagctccctc	660
Db	2807	GATGTGTGTGCTACTGTGTATGACGTTCGATGTGTGTGTGACGCCACGCTTCGCGAGATCCCTC	2748
QY	661	gcagagagcgatgtatgtgcgcgcgcgcacagccgcgcgttcgtttgacgcgctgtcatactc	720
Db	2747	GCAGAGGCGATGATGTGTGCCGCCGCGACACCCGCTGTGTATGACCGCTGCTCATATCTC	2688
QY	721	gcgcgcgcaccttgttcgcacagcaagtytctctgcgcgtctccgcgcgcctcttgacgacgca	780
Db	2687	GCCGCGACACTTGGTGTGCGACGACGAGGTGTCTTGCGCGCTCGCGCCCTGGGCGACGCA	2628
QY	781	ccctgtcccaacagctctctcgtgtgtgtgcccttgtatagacgcgcacagccgtatcgtcgagct	840
Db	2627	CCTGCGCCACAGCTCTCTCGGTGTGTGGCCTTGTGTAGACCCCGCACGCGTATCATCTCGCGCGT	2568
QY	841	gtcgagaggaattgtgtgcgcgcgcgtccgcgcgcgtgtgttcatacagcgcgagcgctgtgtc	900
Db	2567	GTCGAGGCGAGTTGATGTGCGCGGCGGTGCCGGGGGTGTGTCATACAGCGCAGCGCTCTCGTC	2508
QY	901	ctcgacgcgtctcgtctgaatctcaacgttccgcgcacagcagacgttgtgttagacgcgcgga	960
Db	2507	CTCGACGCGTCCGCTGAATTCACACCGTCCGACACCGACGACCTGTCTGTATGACGAGCGCGA	2448
QY	961	agcccccgaagccgcagctgcgtcgcgcgcgcgttcgtgtgtcat	999
Db	2447	ACGCCCGACGCGCACGTCTCCGCGGCTGTGCGGTGAT	2409

[illegible]

QY 181 gaagccacacccgagcgtctccgagagtgctgcgcagaaactctcgtlaagcctcgat 240
 |||||||
 Db 3227 GAAGCCACACTCGGCGGGGCTTCGGCGAGAGTTGGCGAGAACTTCGTAAGCCCTCGAT 3168
 QY 241 cgcgcggcgaggaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300
 |||||||
 Db 3167 CGCGGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3108
 QY 301 gcccgcggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 360
 |||||||
 Db 3107 GCGCGGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3048
 QY 361 gaagacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 420
 |||||||
 Db 3047 GAAGAGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2988
 QY 421 cgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 480
 |||||||
 Db 2987 CGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2928
 QY 481 ctccctgc 540
 |||||||
 Db 2927 CTCCTGCGGCGGCGGCGATGCGCAGCCGCGGAGATTGAGAGAGCCGCGAGAGACTT 2868
 QY 541 ggc 600
 |||||||
 Db 2867 GCGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2808
 QY 601 gatgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 660
 |||||||
 Db 2807 GATGTGTGATCTGTGTAGAGCTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2748
 QY 661 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720
 |||||||
 Db 2747 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2688
 QY 721 ggc 780
 |||||||
 Db 2687 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2628
 QY 781 cgc 840
 |||||||
 Db 2627 CCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2568
 QY 841 gtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 900
 |||||||
 Db 2567 GTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2508
 QY 901 ctgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 960
 |||||||
 Db 2507 CTCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2448
 QY 961 acgc 999
 |||||||
 Db 2447 ACGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2409

RESULT 5
 AF263245 47981 bp DNA BCT 17-SEP-2000
 LOCUS Micromonospora megalomicea subsp. nigra megalomicin biosynthetic
 DEFINITION gene cluster, partial sequence.
 ACCESSION AF263245
 VERSION AF263245.1 GI:10179840
 KEYWORDS
 SOURCE Micromonospora megalomicea subsp. nigra.
 ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Micromonosporinae; Micromonosporaceae;
 Micromonospora.
 REFERENCE 1 (bases 1 to 47981)
 AUTHORS Volchegursky, Y., Hu, Z., Katz, L. and McDaniel, R.
 TITLE Biosynthesis of the anti-parasitic agent megalomicin:

JOURNAL transformation of erythromycin to megalomicin in saccharopolyspora
 MEDLINE erythraea
 20430101 Mol. Microbiol. 37 (4), 752-762 (2000)
 REFERENCE 2 (bases 1 to 47981)
 AUTHORS McDaniel, R. and Volchegursky, Y.
 TITLE Direct Submission
 JOURNAL Submitted (03-MAY-2000) Kosan Biosciences, Inc., 3828 Bay Center
 Place, Hayward, CA 94545, USA
 FEATURES
 source Location/Qualifiers
 1..47981
 /organism="Micromonospora megalomicea subsp. nigra"
 /strain="NRRL3275"
 /sub-species="nigra"
 /db_xref="taxon:136926"
 complement(<1..144)
 /gene="megT"
 complement(<1..144)
 /codon_start=1
 /gene="megT"
 /transl_table=1
 /product="MDP-4-keto-6-deoxyglucose-2,3-dehydratase"
 /protein_id="AAG13906.1"
 /db_xref="GI:10179841"
 /translation="MGDRVNGHATPESTQSARIFLTRHGGPEFATDVIDHMLRAAE
 HRLR"
 928..2061
 /gene="megD1"
 928..2061
 /gene="megD1"
 /codon_start=1
 /transl_table=1
 /product="MDP-4-keto-6-deoxyhexose 3,4-isomerase"
 /protein_id="AAG13907.1"
 /db_xref="GI:10179842"
 /translation="MAVGDRRLGRELGMARGLYWGFGANGDLYSMILSGRDPWTN
 VERLAAGRPYASRAGTWYGDHRTAAEVLADPEFTGPPDAAMWYAHCPAASWA
 GPFREFARTDASVTVADWDLQQRCAVLTELSRLDVLNDFAREVPAVLGTA
 LKGVDPDRLSMTSATRYCLDAQVSPQOLAVTEQLADEIDAVTGGRDAAVLGVA
 BELAANTGNNAVLAETLELAEARLADDEPRTATRYTGVSRTPGVHDERTAASDR
 VGVDPVPGGEVTVAAANRDPEVETDPRDVEDVGGDAETLSRPSGSPRDIIDALV
 ATLATRALRAAPVPLRLSRSGPVLRRRRSPARLSCPVL"
 2072..3382
 /gene="megD1"
 2072..3382
 /gene="megD1"
 /codon_start=1
 /transl_table=1
 /product="MDP-megosamine glycosyltransferase"
 /protein_id="AAG13908.1"
 /db_xref="GI:10179843"
 /translation="MRVFFSMANVSHLFGVPLASAFQAAGHEVAVPALNDDVT
 GAGLTAVPAGDDELVEWHHAHAGODIVEWMRLDWDVDSHTMSMDLLGQOTPTPT
 PALMSPSLIDGWEFCRSMRPMIDVTEPLFAPIAARVYGRPHAMINGPDVATR
 AROSTRLIAHDEVEHREDPLAEMFDWTLRRGGDPHLSTDELYLGQNTDPIEPL
 RIDTGVRTVMRYVYNGSVPAWMLNEPERRYVCLTLGGSSRHGIGQVSTIGMLD
 AIAIDAEFAVATFDQOLVGVSVPANRYTAEPVPMNLPTCAATVHHGGTSMATA
 AIHGVPOIISDADTEVHAKELODLAGISLTPVAGMTAEHLRGATERVLDEPRLGA
 EBRDGMRTDPESPAQVVGICDLDADRAARGHOPRTAEPLHR"
 3462..4634
 /gene="megY"
 3462..4634
 /gene="megY"
 /codon_start=1
 /transl_table=1
 /product="mycarose O-acetyltransferase"
 /protein_id="AAG13909.1"
 /db_xref="GI:10179844"
 /translation="MVTSTINDTTRAPALNSLTGMRFYAAFLVFTTHLSRLIPNSYV
 YADGDAPWOTGTVGSFFLTSGFVLWTSARASDSVSWFRRCVKLFPNHLVTAF
 AAVALFIVTGOAVGEALIPNLLIHAMPALLETSGINPVSMISLACAEFLICPLP
 LFVIGSTRPERLMAAAVFPAAITAVPVVADLLPSSPLIPGLEYSAIQDFELTLP
 AFRSLFETLITLIRLITGRMINVGLPVLFPVFLPFAVSLPFGVAYAISSMMLTP

QY 430 gtacagcactggtgagagaccatgcccagaggtgagcgagcgagcttctcgtgc 489
|||||
Db 8645 gtacagcactggtgagagaccatgcccagaggtgagcgagcgagcttctcgtgc 8704
QY 490 gggcgagatgtagcagcccgagagcttgcagcgagcgagcttgcagcgagc 549
|||||
Db 8705 gggcgagatgtagcagcccgagagcttgcagcgagcgagcttgcagcgagc 8764
QY 550 gaagagcgttcatgagcccgagacactgctcccaagcgagcgagcgagcttgcagc 609
|||||
Db 8765 gaagagcgttcatgagcccgagacactgctcccaagcgagcgagcgagcttgcagc 8824
QY 610 catctgtagagcgttgcagcttgcagcgagcgagcgagcttgcagcgagc 669
|||||
Db 8825 catctgtagagcgttgcagcttgcagcgagcgagcgagcgagcttgcagcgagc 8884
QY 670 gatgatgtagcagcccgagacactgctcccaagcgagcgagcgagcttgcagc 729
|||||
Db 8885 gatgatgtagcagcccgagacactgctcccaagcgagcgagcgagcttgcagc 8944
QY 730 ctgtgtagcagcgagcttgcagcgagcgagcgagcgagcgagcgagcgagc 789
|||||
Db 8945 ctgtgtagcagcgagcttgcagcgagcgagcgagcgagcgagcgagcgagc 9004
QY 790 cagctcctgagcttgcagcgagcgagcgagcgagcgagcgagcgagcgagc 849
|||||
Db 9005 cagctcctgagcttgcagcgagcgagcgagcgagcgagcgagcgagcgagc 9064
QY 850 gttcagtcagcgagcttgcagcgagcgagcgagcgagcgagcgagcgagcgagc 909
|||||
Db 9065 gttcagtcagcgagcttgcagcgagcgagcgagcgagcgagcgagcgagcgagc 9124
QY 910 tccgctgtagcagcgagcttgcagcgagcgagcgagcgagcgagcgagcgagc 969
|||||
Db 9125 tccgctgtagcagcgagcttgcagcgagcgagcgagcgagcgagcgagcgagc 9184
QY 970 cgcgagctgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 996
|||||
Db 9185 cgcgagctgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagc 9211

RESULT 6

AF147704 5908 bp DNA BCT 08-FEB-2000
LOCUS AF147704

DEFINITION Streptomyces fradiae NDP-hexose 2,3-enoyl reductase TylicII
(tylicII), NDP-hexose 4-ketoreductase TylicIV (tylicIV), NDP-hexose
3-C-methyltransferase TylicIII (tylicIII), mycarosyl transferase
TylicV (tylicV), and NDP-hexose 3,5- (or-5-) epimerase TylicVII
(tylicVII) genes, complete cds.

ACCESSION AF147704.1 GI:5305790

VERSION AF147704
KEYWORDS Streptomyces fradiae.
SOURCE Streptomyces fradiae.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Streptomyces; Streptomycetaceae; Streptomyces.

REFERENCE 1 (bases 1 to 5908)
AUTHORS Bate,N., Butler,A.R., Smith,I.P. and Cundliffe,E.
TITLE The mycarose-biosynthetic genes of Streptomyces fradiae, producer
of tylosin

JOURNAL Microbiology 146 (Pt 1), 139-146 (2000)

MEDLINE 20121747
PUBMED 10658660

REFERENCE 2 (bases 1 to 5908)
AUTHORS Bate,N. and Cundliffe,E.

TITLE Direct Submission
JOURNAL Submitted (30-APR-1999) Biochemistry, University of Leicester,
University Road, Leicester LE1 7RH, UK

FEATURES
source 1. 5908
Location/Qualifiers

/organism="Streptomyces fradiae"
/strain="759235"
/db_xref="taxon:1906"

gene complement(80..1069)
/gene="tylicII"
/note="orf10*"
CDS complement(80..1069)
/gene="tylicII"
/function="biosynthesis of NDP-mycarose during tylosin
production"

/codon_start=1
/transl_table=11
/product="NDP-hexose 2,3-enoyl reductase TylicII"
/protein_id="AAD41821.1"
/db_xref="GI:5305791"

/translation="MSGMTVOLGRATVLSRLMCTVNSGHPDQAIRLMDALDR
GVNCDVADIVGMRLYKHTBELVGRMWSGSRDDVYATRVGPMDSRVDRGLSA
RHVNSCEASLRLRLGVDHIDLYOMHMDRTVMQMDQVLSGKRYTGSNFA
GWHLAGOESARSGSLGVSBQCLNLAIRAELEVLPAARATGIVGAVSPGLGL
LSGALRKLAEGTAVKSGGSRADRTIPALBDITARYERERCARVGRDPAEYGLMLSRP
GVSGAVIGRTTGHVLSALRAYELESEHEHLEALFPVSGGEVPAWON"
gene complement(1066..2109)
/gene="tylicIV"
/note="orf9*"

CDS complement(1066..2109)
/gene="tylicIV"
/function="biosynthesis of NDP-mycarose during tylosin
production"

/codon_start=1
/transl_table=11
/product="NDP-hexose 4-ketoreductase TylicIV"
/protein_id="AAD41822.1"
/db_xref="GI:5305792"

/translation="MSMPLVPGPARDRRPLVVLGASGFLGSAAVAELARRPRLR
VARBPSPVAGVAELIEVRRDLSPEVTEAEDAVVHLVAHTGERSVRAAGER
SEHINAGLEBELAEALRPDPADATAPAVVLGSLQAOEMEARHTEGTYAAKLAER
VLRADAEAGVAGVRLRTYIGRSPGLTSGPRGVLAAYAGALAGDPTTMHDSVE
RLDLYRDAAATFTTALHADOLRGKHWVYGGRRHRLDRVGTVAALAAEHTGRPV
PVYVDPPEYAEVCDPRTDSDPSAFRAVYGNRPRAEDGIAAIAVAAGDSPEP
EGAGKRGGGERK"
gene complement(2106..3362)
/gene="tylicIII"
/note="orf8*"

CDS complement(2106..3362)
/gene="tylicIII"
/function="biosynthesis of NDP-mycarose during tylosin
production"

/codon_start=1
/transl_table=11
/product="NDP-hexose 3-C-methyltransferase TylicIII"
/protein_id="AAD41823.1"
/db_xref="GI:5305793"

/translation="MPAVPRDQMTISACRVGNRELLFVLDEGALGVPPRRRE
TVPSIPLELVKCSPOGGLVLRHTPDGLMAGEGYSGIRPEMIGHLRRKVAIR
ELVDLGGDDLVLDINDATLLKAYEGRGLIDPSGGRFELVPPHAEILAEYFSR
DVETAPGRRRARVITSTIMFYDLPPLAFMFDVHVDLADCTIWMEOGYLXLALEAD
AIDIVCHEHLEYALDQITMAERACGLVLRLELDYVGGSLCYTLARRSSHPRDEA
GPARIARTEAKINTMAPEEFARVEVRRDLRDLRSRAGRLTLVGASTKGN
VILQYQIGERDLPICGEVSPKAGFTFGTGTIPVSEEDAKAMPDQLLVLPWYRE
GEVERERDPLAGGRVLPPLRLDYV"
gene 3575..4960
/gene="tylicV"
/note="orf7*"

CDS 3575..4960
/gene="tylicV"
/function="catalyses the incorporation of mycarose during
tylosin production"
/note="glycosyltransferase"

/codon_start=1
/transl_table=11
/product="mycarosyl transferase TylicV"
/protein_id="AAD41824.1"
/db_xref="GI:5305794"
/translation="NAGLRPGAGVPPGTPMPSPKHCYSAISRARLPYAPGTPPL
GRRRAPDGGSCRMHAHAFITLLPAGHVPPTLGVAEELARHRTYALPEMDADRAV
RGARAVTYPLDRERRADMPKESDEYTDGEEFLKYLEWLLDTTADTLPLESAFA


```
gene
CDS
/translation="MCVDRDGRGAGVTADLTARCARAAPEVADVSDGAAEREAEER
VAETVGVVDLLVNNAGISAGRLDTSVEDWQRTGLVNGVTHGCRLLGRMAERGO
GGHIVTAAAFQPTPAVPAAYATSKAAVLMSECRAEFAEAGVGVSVCCGFVTS
PASAMFAGVPRLEDERLRALFAGRCSSAEKVAAVLRVSARDAVVVTAEARLSRL
MSHFPRLRRAAARMDPPS"
931..2169
/gene="aveB1"
/gene="aveB1"
931..2169
/feature="transfer of oleandrose onto polyketide-derived
acylons"
/codon_start=1
/translation="MSDIFLMSAPFPGVHPSPSLAVALAEELVHGHVTEVGAEMADA
VRSVGDPLRYESAPFQDMYRLMTEAPNALIPMLTYDEGMSKLRSEEHKGDVVDL
VADTATSLANGRVLAASWSRPMATYIPLEPASGRESTMQSVLDPDSAOVSAPPRRS
EOMELFGALVPRLAELLYSRKITEPVDFLSGPDENLVCLPRAQYAGDTEDERS
AEVGPCLGRKRGENTPPGSGHPVLLISLGVFNRLSEFFETVRAFTDVPHVVIS
LKGVPDLRLPPLPVEVHRWPHHVALEHARALVTHGCTGSVMEALAGCPVLVMP
LSRDAQVGRRIAEGLGRMVQPEEVATTLRRHVLTISDDAIFRQVNMQRATVEA
GGAIRADETERFLRTRRH"
complement(2075..3142)
/gene="aveB11"
complement(2075..3142)
/feature="function in oleandrose biosynthesis"
/codon_start=1
/translation="dtdp-glucose 4,6-dehydratase"
/protein_id="BA84593.1"
/db_xref="GI:5921157"
/translation="MTGAGFTGSHFRRLTGAPYAPFTGAEVYVLDKITYAGRLNTL
AYLGSPLIFVHGDCDPLVADLMDGSDMYVHPFAESHVDRSVADAEPRTNVLG
HTLLRATDAADRYTISTDEVTGSDSGTEADPLEPNSPYSASASDLLARS
FRTHTGLPVIITRCSNNYGHQPEKLIPIREYTHLNGTKVPLYGGEVNRMLHARD
HCRGIALVABERGPEIYHIGGTELSNRELRALDLGVMSWEPEYTDKRGHNR
YSLDISKISAEGLYARVPPEEGLAQTVQVYENRTLMPELTARPELPVSDGASMET
ARSRLPACRPPRPMPASA"
complement(3139..4038)
/gene="aveB111"
complement(3139..4038)
/feature="function in oleandrose biosynthesis"
/codon_start=1
/translation="glucose-1-phosphate thymidyltransferase"
/protein_id="BA84594.1"
/db_xref="GI:5921158"
/translation="MKGIYLAGTGSRLYPLTRALSQKLLPYDKPMIYYPYSLMIG
GIKDLIVISSPHLEQFRRLDGBSRLGNTIYAAQORGGIAEAPLIGADIGOOOV
SLVLDNIFHGMGFSHLRSHTRDVGCVLFGVAVDPRYGVGASGKLLSVBEK
PPAPRSNLATIGLYXNDVYIYVARGIRSSARGELEITPYNRAYIAGEGARLYDLCRG
ETWLDAGTDSLHMAQIYVLEKROGVRIACLEIAFRMLGLIDADDCTLRGVELTGS
4164..5195
/gene="aveB1V"
4164..5195
/gene="aveB1V"
/feature="function in oleandrose biosynthesis"
/codon_start=1
/translation="dtdp-4-keto-6-deoxy-L-hexose 4-reductase"
/protein_id="BA84595.1"
/db_xref="GI:5921159"
/translation="MGRSVCPRPPIGLKSMITGMDPLVVLVGLASGYIGSAVAA
ELARWVLLRLVARRPVPGGAETERTADLTAAESVALAVDADAVIHLVART
OGAAMAAESDPAERVNVGVMDVVAALRSRGRRGPPVYVAGSVYOVGRGRVGD
SEDEVTAYAROKLDAERTLKSAVEGALSLPLPYVAGPGPGNGVQVQVAVL
RLADEALTVMNGSVYERDLVSEVDAQFAVGCIALHADLRHMLTSGSGRYTPVHL
PGAIAGVSAKTRGPAVPTAVDPPAMAADPHGIVYDSSAFRAVTKGRPLSLDQEG
```

```
gene
CDS
/translation="LDMHMAVAV"
complement(5249..5926)
/gene="aveB1V"
complement(5249..5926)
/feature="function in oleandrose biosynthesis"
/codon_start=1
/translation="dtdp-4-keto-6-deoxyhexose 3,5-epimerase"
/protein_id="BA84596.1"
/db_xref="GI:5921160"
/translation="KTLGRPRSSADRPAPGARTAGVVRRLVGEAGVEFPPTV
PFDERLEVTYQDEPLSEAVGRREPTAQCSVSRRGVNRCVHFTAPRGQAKVYHC
ARGRALDVLVDLRTGSPITFGQMDSVLLDGERRSVYLPVGVHAFVALLDDQAMVLM
SSGYVQNHALSPEDDPLALPLGHLHGRAPILSERDRAPTLQOALRGMLPEYRAS
RGIDEKL"
complement(5988..7397)
/gene="aveB1V"
complement(5988..7397)
/feature="function in oleandrose biosynthesis"
/codon_start=1
/translation="dtdp-4-keto-6-deoxy-L-hexose 2,3-dehydratase"
/protein_id="BA84597.1"
/db_xref="GI:5921161"
/translation="MSVRADADHTERTAHRAARRPARVPRLRGRHRRRTSLDA
FPGWTRRSAGHRRFYERIPFHGMAMSPHCTGNLAHSGRFSFEGCIHVGQDOP
PEROQPIHOPETIGITIAKRDYVHLKMAKMBGNINLVQSLPTVQAKRSNTK
VHGGAUVKYLEYETOPRRATVVVDVLOSHEGAMFHKRERNVTEDEVDPLDDERW
LTLGOIGELMHNDNLVMDARTVLAICLPPEPEPALHSDADELSWYABERSHSYHA
RRVPLAGIPGHTTGAESIANHADRYFRVAVAEASNRVSAWMTOPLEPCSGHTAF
LRRIGGVNHLAHGVEGGPLDTLEGTQYTPRNVAHLGAPRRLDVLGEAP
DRIRAGVAHSEEGRHLNARVLEYEADSEAPNDPPRGYKWCIPGQTLRLRGYR
VWQARTLSLITTRAVEL"
complement(7394..8167)
/gene="aveB1V"
complement(7394..8167)
/feature="function in oleandrose biosynthesis"
/codon_start=1
/translation="dtdp-6-deoxy-L-hexose 3-O-methyltransferase"
/protein_id="BA84598.1"
/db_xref="GI:5921162"
/translation="MTEREFTPRIYHESBOBARAEOYTKLWDSPIPPXYLIDNL
SVYMRNOLADLSDALYRMLPEVDPGMEGVLAGRLATLTALRSIYEPYNSLR
VIGFPTFTFPDIDENDEYSTASVGRFVAPGVEVNLQOVLAANEPRYCHTQSF
VVOGDVRETVPOYLAENPHTVIALAFDIDLTPRELVDVTPHLLTRESIIAFDELT
HKMPGETRALSBAFGDLHAPLROLPGRPPVIYMTWGA"
complement(8254..9297)
/gene="aveB1V11"
complement(8254..9297)
/feature="function in oleandrose biosynthesis"
/codon_start=1
/translation="dtdp-4-keto-6-deoxy-L-hexose 2,3-reductase"
/protein_id="BA84599.1"
/db_xref="GI:5921163"
/translation="MPTTAEPAPVOSDSSNAPLTELGRRLAISRLATGVNIG
RVEEPARRLMHDLAQGITLPTANTGVHKKGYTDEVTIGRLADRPAREQVYLA
TRVGDPMGSPNDHGLSVRNIVAACDASLRRLTMDIYOLHNLDRAGMDEVOQAM
DLITOGKRVYVSSNFGAMDIAQOEARRRNALGLASEOCVUNLVTHADELVYPA
ASAYGVLYWSEPLGGLGVLRTRRENTAVKSAQGRAVELNHNRTITAYEDVCA
DHGIDPRANHMAVLSRPGVTLYVIGPRIEQVNDGALNHLRTPLREPVYLARELEFP
VERGGSAPRANLIS"
BASE COUNT 1341 a 3029 c 3601 g 1559 t
ORIGIN
Query Match 51.2%; Score 511.6; DB 1; Length 9530;
Best Local Similarity 70.6%; Pred. No. 1.6e-46;
```


gene 3462..4634
/gene="megy"
CDS 3462..4634
/gene="megy"
/codon_start=1
/transl_table=11
/product="mycarnose O-acyltransferase"
/protein_id="AAG1909.1"
/db_xref="GI:10179844"
translation="WVTSNLDITTAAPALNSLTGMREVAELVEFTHVLSRLIPNSVY
YADGLDAWMQTGVRGSEFFILSGVFJTMRSARASDSVMSFMRRCVCKLPNNHLVTAFA
AAVFLVGTGAGVSEALIPMLLIHAFEPALLETISGIPVMSLACEAFYCEPLF
LEWISGIFPERMAMAAVFAFANVPAVPAVLDDLPSSPGLTPIGEAIDMVLTPPP
ATRSFELIGITLIRILITGRMINVGLIPALLPPLPPFVSLPIPGVAILSSSMILIP
LVLIITASATNDLOOKRTFPMNRNVMVHGDVSPALYMHPIVLYIYGCDLIGFSTDEA
PLGLALEPIIPELAVLSLWLTIRFELLPMNRMARPASARKRPATEPQTSRR"
4651..5775
/gene="megdII"
CDS 4651..5775
/gene="megdII"
/codon_start=1
/transl_table=11
/product="UDP-3-keto-6-deoxyhexose 3-aminotransaminase"
/protein_id="AAG1910.1"
/db_xref="GI:10179845"
translation="MTTYVMSYLLFYERERADILDAAQKFAAGSLILGSGVNFETE
YARYGICHGCVGDNGTNAVKTALESVSGSDGDDVMVNSNATFVAIDIEIGAPVAF
VDVREDIEIMOTDLVEAAVPRTRAIYVPHLYGQCVDMPTALRELDARGLKIVEDCAO
AHGARDRIAGTMSDAAAFSEFYPTKVLGATGDDGAYVYNDDETARALRLRYGME
VYTRTGRHNSRLDEVOAETLRRLTRLDAYAGRAVADORYDGLADLSDSGLEL
PVYDNGHNVAVVYVVRHPRRDEILIRLDDYDISLNIYPMVHTMTGFAHLGVASG
SLPVERLAGEFELSIPMPSLPDLDDRVIEAVREVITGL"
5822..6595
/gene="megdIII"
CDS 5822..6595
/gene="megdIII"
/codon_start=1
/transl_table=11
/product="daunosaminyl-N,N-dimethyltransferase"
/protein_id="AAG1912.1"
/db_xref="GI:10179847"
translation="MPSNSTSTSDVAPYERADIDHEFYHGKRGYRAEADALVEYA
RKHPOATMLDLYACGSGSHLVELADSEREVGDLSAMATLTAARDPREFHOGCM
RFSQDRFDVYTCFESSTGYLVDEAELDRVAYNALAGHLAEGTILVDEPPMFPETFP
GNVGVLDRTSGGRTISRMSHTVPAGLPDRTRSMTHYTVGSPGAGIEHTFEVHMVL
FARAAYEQAFRAGLSCSYGVHDLFSPGLFVGVAAPGR"
6592..7197
/gene="megdIV"
CDS 6592..7197
/gene="megdIV"
/codon_start=1
/transl_table=11
/product="UDP-4-keto-6-deoxyhexose 3,5-epimerase"
/protein_id="AAG1911.1"
/db_xref="GI:10179846"
translation="MRVELGIEGVFTFTPOTFADRGVGTATAOEDVFAAIGRPLF
PVASQSTRSRGVRGCHFTPMGSAKKYYKACRGAMPADVIRGSGPTFGAEVY
ELDSASWYTLRPGMGHLVSEDDTILVYLSAGYVPDKERAHPLDDELALPIPA
DULDSMRDRAPLPLREARDQGIIPDYAACRAAHNVRT"
7220..8206
/gene="megV"
CDS 7220..8206
/gene="megV"
/codon_start=1
/transl_table=11
/product="UDP-4-keto-6-deoxyhexose 4-ketoreductase"
/protein_id="AAG13913.1"
/db_xref="GI:10179848"
translation="MWVIGASGFLGSANTHALADIPYRVLVAREVYVPSGAVDAE
THRDVLPSCALAEVADARAVFFPPAAQIRGTSWRISSEDDVYAEKRVNGVRLDAV
LSRSHAVVVPFGSNTDVGKVTAGRYIDGSEODHPGKGVYDROKHTEQILKEVTAAG
AIRATSLILPVPFVPAAGTADDRGVASTYMLRALYTOQPLIMHMDGVYRELLVYDPA
ARAFTALDADLAGRHFLLTGRSPVLCGEVQAVSRVSAVHTGDEPVRVSPPA

[illegible]

	D	46408	CAGTGCACACTCTCGGCGGGGTGTCGCCCCGAGAGTCTTGACAGAACACTGTAGCCCTCAC	46349
QY	241	cgcgcgagcgcgaaggagcgcaaacagcacctctgcacagcgccctcgcgcgaattcaaccgggt	300	
D	46348	CGCCGACCGGACATGTCCTCTCGCCTGCCACACGCCCCCTGGAGCTGGGGGCCCGACCG--	46289	
QY	301	gcccgcgccagcttctccagcgctccgcgttgagcaagcgccgcgttcagcggcgaaacgc	360	
D	46290	-----TCGCGCCGACACAGACCCTGGCAAGCGGCTGGGCTTGGC	46253	
QY	361	gaagaagccgaagcccgttaagcctctgcgcgcgcgcgcgcgcacatcccaactcgcgcgtgcgc	420	
D	46252	GAAAGACCCCGACCCCGAAGCCTGCGCGCGGGCAGAAGACTCTCAATTGCGGATGGCGCGA	46193	
QY	421	cgcgcagtcttacagcagcttgtgtggagacacatgccccaaggagtgcgcgcgcgcgtt	480	
D	46192	CGTAGGTCCTACCGACACTGGTGGGACACAGGCGGAAGCGGTGACGCGGACCGGCTG	46133	
QY	481	ctctctgcgcgcgcgcgatctgtgcacgcgcgcgcgaatttcgcgcgcgcgcgcgcgcgcgc	540	
D	46132	CTCTTGGGCGGACACAGTGTGCCATTCGGGGAAGCCCGACGACCCGATACACACTT	46073	
QY	541	gcgcctgc	600	
D	46072	TCCGGCGGCGACGAGGGCGTCCACGCGCTGCCAGACTCTGTCTCCAGCGGCTCACACCGGCTC	46013	
QY	601	gatatgtcatctctgtgttagacgttcgatctgtctgcacgcgcgcgcgcgcgcgcgcgcgc	660	
D	46012	CACCCGGGGGACAGGTGAAGGACGTGCAGCTGTGCACACCGAAGACCGCCAGGAGACCTTC	45953	
QY	661	gcgaagaagcgaataigtgc	720	
D	45952	ACAAGAGGCGATATCTGCGCGGGGAGAGAGCGCGCCCGCCGACTTCCCGCCGGTGG	45893	
QY	721	gcgcgcgcaccttgctgc	780	
D	45892	GACACCCACCGCTGACCGACGACGAGAGCGGTCTCTCCGCGCACCGGCTGTCCCGGACCA	45833	
QY	781	cctgcaccacagctcctcgttgtgtgccttttagagcgcgcgcgcgcgcgcgcgcgcgcgcgc	840	
D	45832	CGGCGACGACGACTCMTCGCGGCACTGGGCGACATGGTGCAGCGAGATGCTGTGCGCGGT	45773	
QY	841	gtcgagcgcagtatactgcgcgcgtgccgggcgtgtgttcatacgaagcgcgcgcgcgcgcgc	900	
D	45772	GTCGAGGCACTTACACGCCCGCGAGAGGGCGTGTGCAGCAAGCTCATGCTCGCGCS--	45713	
QY	901	ctcgagc	960	
D	45714	-TCGCTCTGCGCGCGGCGAGGGCTGCTCCTTACCAACCGACCGCGGTGTGACGAGGGCGCA	45656	
QY	961	acgcgccgaagccgaactgc	997	
D	45655	CCGGCGAGCCGGACGTGGGTGGCTGTGCTGCTGCATTC	45619	
RESULT	9			
LOCUS	SCD19/c			
DEFINITION	SCD19 bp DNA	BCT	18-OCT-2000	
ACCESSION	AL392149			
VERSION	AL392149.1	GI:10129746		
KEYWORDS	ABC transport system ATP-binding protein; ABC transport system integral membrane protein; aminotransferase; DNA invertase; integral membrane protein; integrase; nucleosidytransterase; oxidoreductase; oxidoreductase lipase; tetr family transcriptional regulatory protein; two component system response regulator; two component system sensor kinase.			
SOURCE	Streptomyces coelicolor A3(2).			
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces			
REFERENCE	1 (bases 1 to 22991)			
AUTHORS	Redenbach,M., Kleser,H.M., Denapalte,D., Eichner,A., Cullum,J.,			

TITLE	Kinashi,H. and Hopwood,D.A.
JOURNAL	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome
MEDLINE	Mol. Microbiol. 21 (1), 77-96 (1996)
REFERENCE	97000351
AUTHORS	2. (bases 1 to 22991) Seeger,K.J. and Harris,D.
JOURNAL	unpublished
REFERENCE	3. (bases 1 to 22991) Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A. Direct Submission
AUTHORS	Submitted (11-sep-2000) streptomycetes coelicolor sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrelle@sanger.ac.uk Cosmids supplied by Prof. David A. Hopwood, [3] John Innes Centre, Norwich Research Park, Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT	Notes: Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics Details of S. coelicolor sequencing at the Sanger Centre are available on the World Wide Web. (URL: http://www.sanger.ac.uk/Projects/S-coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS. Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot program of Bibb et al., Gene 30:157-66(1984) as implemented at http://www.nih.go.jp/ jun/cgi-bin/frameplot.pl . CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or attt) which is preceded by an upstream ribosome binding site sequence (optionally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon. IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid D19. Location/Qualifiers 1..22991 /organism="Streptomyces coelicolor A3(2)" /strain="A3(2)" /db_xref="taxon:100226" /clone="cosmid d19" complement(1..358) /gene="SCD19.01c" complement(<1..358) /gene="SCD19.01c" /note="SCD19.01c, unknown(fragment), len: >119 aa. Contains two TTA leucine codons, possible targets for bldA regulation" /codon_start=1 /trans_table=1 /product="hypothetical protein SCD19_01c (fragment)" /protein_id="CAC08302.1" /db_xref="GI:10129747" /translation="MORTYGLKTELVDRLNGPPEGVTDLEAAQLVLYHTELENNCT GGGEKLNEEMAEALRTLKFLLRLLKLDLKAAPDDFOSFRKYWIREMGGGGAGAKRRR SYDGLEYEVPVRERIDEM" 1..98 /note="nominal overlap with Streptomyces coelicolor cosmid SCD12A" complement(35..37) /gene="none"
FEATURES	
source	
gene	
CDS	
misc_feature	
gene	


```

misc-feature      complement(35. .37)
                  /gene="none"
                  /note="TTA leucine codon. Possible target for b1dA
                  regulation"
                  /label="TTA
138. .143
misc-feature      /note="Drai site matched TTAAA"
                  /label="Drai
complement(140. .142)
/gene="none"
/note="TTA leucine codon. Possible target for b1dA
regulation"
/label="TTA
gene             complement(140. .142)
                  /gene="none"
gene             complement(603. .1031)
                  /gene="SCD19.02c"
CDS              complement(603. .1031)
                  /gene="SCD19.02c"
                  /note="SCD19.02c, unknown, len: 142 aa, High content in
                  alanine, arginine and leucine amino acid residues"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein SCD19.02c"
                  /protein_id="CAC08303.1"
                  /db_xref="GI:10129748"
                  /translation="MTNHQAEAEERIRSRAMORLLAGVDPGLKCDKSLCTAGVPR
                  ALTRYTPHLKAEFRDRTAADAGQDPTRLAQRERLKAEVATLERLRERLRKKNVEIDA
                  LKEFVYALSRITAAODATSLRRELEAQRERLRVPR"
                  complement(1028. .3199)
                  /gene="SCD19.03c"
                  complement(1028. .3199)
                  /gene="SCD19.03c"
                  /note="SCD19.03c, hypothetical protein, len: 723 aa;
                  similar to some plasmid derived hypothetical proteins,
                  e.g. TR:085873 (EMBL:AF079317) Sphingomonas
                  atomaticivorans hypothetical 81.3 kDa protein (plasmid
                  pML1) 741 aa; fasta scores: opt: 421 z-score: 478.1 E():
                  4e-19; 29.7% identity in 758 aa overlap"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein"
                  /protein_id="CAC08304.1"
                  /db_xref="GI:10129749"
                  /translation="MTTAASSPALAGRPFGHGLPIETAGLRREPGRPFVDDQDW
                  DITGLADAVVMSTHRRILDTAIIINPRKQVARELMRLAPLHPDVAVLDPQAFRP
                  LPNSLMSLKLALMFNHLTAVGITSLSQVRQHDADAYLAASRSATDPDLSPAT
                  TVAMVRIPOFLVLAELISDQYGFAPMGSRADVTGVSDEKRRVPPDTLLRP
                  LLANCLYLETIGPLVPEPAAARADQHAASRSRLVTEVDGALREVEERRETCIP
                  ASRATVATYQRLKRGMDPGDPLHMSHPIVYQAGANGHRDDETLRLPELERNYSE
                  CGIQQPCWDAAPVRPDDGTIPVPMALPMARHOLDMAYVVTSAIFLTSALSGMSS
                  ELAELTSGKROEERTGGTTRVLRKIGEGTGEDAMVTEDEVQSAIAIAELT
                  GAPGERLPAKSSNSGRYRTALSWNVNDEYGRALIEPLDGPVPRALRLTAAI
                  AORPHGLMAVHLKHLISVATAGYAAARAGAAFAAEAAEAEHEERLRTVLAAL
                  YORGLIPSGOGARBDVAARFSDOALERHDAGGVYVITDRRGRVRLKAKATHTGVC
                  NYCWSPSKALCIKLACTPPADDEPLIGMCDSCRQATHHSHRQIMADHDNTQAV
                  FLGNPRLSPERARQNAFERATRIVAETDAAGHPVEELHS"
                  complement(1042. .1045)
                  /gene="SCD19.03c"
                  complement(3196. .4911)
                  /gene="SCD19.04c"
                  complement(3196. .4911)
                  /gene="SCD19.04c"
                  complement(3196. .4911)
                  /gene="SCD19.04c"
                  /note="SCD19.04c, hypothetical protein, len: 571 aa;
                  similar to some plasmid derived hypothetical proteins,
                  e.g. TR:085872 (EMBL:AF079317) Sphingomonas
                  atomaticivorans hypothetical 65.3 kDa protein (plasmid
                  pML1) 595 aa; fasta scores: opt: 344 z-score: 399.6 E():
                  9.4e-15; 26.7% identity in 561 aa overlap. Contains two
                  TTA leucine codons, possible targets for b1dA regulation"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein"

```

```

/protein_id="CAC08305.1"
/db_xref="GI:10129750"
/translation="MPAEIOQPLAVFLTPDGVSHHGVLDLIPNOQLAVDLAVGLVAA
THRHPGIRTSYARQDMTMMRMARADLAASGEFGIADITPATLQVYITCDYHRRR
IRVYLAPFETAGGIDPGICRHLAGRRIRKVERNPQYISGBERLBSACTNOIAA
ARRTORALEAARGADPSSHGVTEDNLMLWHLHAPARARAERFEGACVDOGH
AAVEALFPTSETAEFAYLTFEMRGTIVPDGILVDLIDITRSANTALLSYRKRTG
NEALNLPDRAVRLDRLWLESHAQRLRDHADLDRLVIYVCGDLGRGQRIFFDRPSQ
ROBRAMESSGVIGDGGOPLPVHGGRVATYHRRRTAMTGRTTIDPMSARVEGD
YLSHHPADLDLEGTIEQAGDVRKRAAPVYSSSEDAARAFADPPRIVEDAGLDA
AVSALISGEODHFNACASPLNGPRAPATILCPARWVLLICPLAFARHLPNLRL
KEYFSRQAOQMTTGQFLRTFPTARLDVLPFRGPAIEATROSALTATFLPLHL
EQQPQ"
RBS              complement(3210. .3214)
                  /gene="SCD19.04c"
gene             complement(3979. .3981)
                  /gene="none"
misc-feature      complement(3979. .3981)
                  /gene="none"
                  /note="TTA leucine codon. Possible target for b1dA
                  regulation"
                  /label="TTA
gene             complement(3991. .3993)
                  /gene="none"
misc-feature      complement(3991. .3993)
                  /gene="none"
                  /note="TTA leucine codon. Possible target for b1dA
                  regulation"
                  /label="TTA
gene             complement(4911. .6263)
                  /gene="SCD19.05c"
CDS              complement(4911. .6263)
                  /gene="SCD19.05c"
                  /note="SCD19.05c, possible integrase, len: 475 aa; highly
                  similar to TR:09RAJ3 (EMBL:AT250372) Mycobacterium sp. GP1
                  putative integrase InIM, 451 aa; fasta scores: opt: 2152
                  z-score: 2497.3 E(): 0; 72.8% identity in 441 aa overlap"
                  /codon_start=1
Query Match      37.4%; Score 374; DB 3; Length 22991;
Best Local Similarity 63.5%; Pred. No. 3.1e-32;
Matches 610; Conservative 0; Mismatches 335; Indels 15; Gaps 2;
QY 42 cggggaagatcgtccagatcggagacagcgctcgtccaggtatcgcgagcgt 101
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7999 cggggaagatcgtccagatcggagacagcgctcgtccaggtatcgcgagcgt 7940
QY 102 tcagcgcgagatcgaagctgtcgggggttcgcggggccgagatcgagcgcgatgcgg 161
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7939 GCAAGCGCGAGTCCAGCTGCTCTGCTGCGGGGCGGACAAATCGGCGCGGTCACGCGCG 7880
QY 162 gccgggacagaccatcgcagcccccacccctcggccgggtcttcgcgaggttcgcgaga 221
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7879 GCGGGGAGACAGCAGCAGCGCGCCCTCGCGGGCTCAGGCGCTGTTTGCAGCA 7820
QY 222 actctcgttagcctcgtatcgccggcgagggagcgcaacagcactcgcgagcgct 261
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7819 GGTCTCTGTACGCGCTGATGTGCTCGCGCTTGTGCTTGTAGCGCATCGGACGCGC 7760
QY 282 ggcgcgagcttcacgcggtgcccggcgagctcttcacgcgcttcgctgagagcgcg 341
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7759 GGCCTCTGCGCCGCGCGCTGTGTGTGCTCTTGTGATCAGCGCGCCAGCAGACCGC 7700
QY 342 cgtgcagcgcgacgaagcgaaagcgacgcgcgttagagcttcgcgcgcgagcagcct 401
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7699 CGTGCACGGCGACACACGGGATATACCCCGAGCCGTATGCTCCCGCGCGGATCATCCT 7640
QY 402 ccagctgcgctgcgagccgacgaggtgtacaggaactgtgaggaacatgcccaagg 461
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7639 CCAATCTCGCGCGCGCTCGGCAAGTTTACAGCATCTCGTTCACAGCCGATG 7580
QY 462 agcgg---cggcgggcggttctctcgcgcgcgagcgatgtccagcccggaagtgc 518
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


gene
CDS
/product="hypothetical protein SCD12A.03c"
/protein_id="CAB93399.1"
/db_xref="GI:8249973"
/translation="WRTAVRRRAATVLTSTAALLAFAFAEAAGVIGDGGITLL"
complement(1987..4590)
/gene="SCD12A.04c"
complement(1987..4590)
/gene="SCD12A.04c"
complement(1987..4590)
/note="SCD12A.04c"
similar to TR:053874 (EMBL:AL022004) Mycobacterium
tuberculosis hypothetical 79.7 kd protein (fragment)
MTV043.55c aa; fasta scores: opt: 651 z-score: 618.2 E():
5.7e-27; 36.1% identity in 886 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein SCD12A.03c"
/protein_id="CAB93395.1"
/db_xref="GI:8249974"
/translation="MSTEERKSAAPRSALAEALRVDDVSLAALRSRDLTPVTDLT
LTAPAGRSASVVRALREDFRLPOTAEALLAVDDPASVSGLLALMGDDGDDPVA
LRPAALLREQALVWGADRLRLVTRTAELLAPSPQHSPGGLGIVREATAGCSPP
TODTLLAAGLPSYHDVAVSVALSGLEFDRRRMAALLAEIPWESREVLDRVWSPPG
OYTPDPAHLRLDLRGLLPLTPAGTVVLPREVALHLRAGAHAAPEEPVQEAAT
HPPQVDAATGAQALAAATATDELKEMDEGSPVTLFAGLSVDLKRITAAIDVPEP
VAGFVEIATAGGLIASDGEAEERYAAVPADEMRLEPAPRRARLGSMTLTARTPG
VAAFDADRTLSALGPNLDRSAAPEVVRVALLAGLPEASPAVSATVRLAWEPP
LRGPOQATGACGAREDDLSRIARKWTLSSBELGCTGRALAAPGALIGAEAPR
PATNADTVGGPGEDKLPVHHHRTPEVPAAPTPAERAAATVAAHLAPLEPELDHV
LQADLTAVAPGPERGLADVLADVDESFGATVYRFPGSVRALMDGOSADHLH
AFIAHSRTPEPOPPLYLIDVYARRHGRLRGAASAVYRCDDDTDEILADKRAAL
GLRRLAPVLAADQAPPAALLDLGLRAGPAAEASAGVYLIRADDSHRTPAPEV
LATGPAPDDTLLAARAVRAGDLASTPRKPGCDGDEGSGSCMGPIPRCAEAT
LATMPAAVTEGALWIGTVNAEGASQNRVAPIRVEGGSVTAYDHTADEVTPYLRRV
TGVAAELADGAG"
4640..5740
/gene="SCD12A.05"
4640..5740
/gene="SCD12A.05"
/note="SCD12A.05, possible integral membrane protein, len: 366 aa. Contains possible hydrophobic membrane spanning regions at C-terminal domain. High content in alanine, arginine and glycine amino acid residues"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAB93396.1"
/db_xref="GI:8249975"
/translation="MRCRRPGGGRGTGVRREGTGHGVSQRTAPEGNVGIESQVVF
EYLSVGVGAOQRLPSATRMKLVSLENEIDRRIRAKTVDSPAARIRIDRLDDPAD
IVTAAGVSGVRQOQAPSPATVPEQRDVGPPDAARRGRSPAEETPAAPPAASPPIL
ASARELGVGGQDDPMWRVDDGPFAGGGLGDSVGVGVEIPILRRKPGVADEEER
TEPAPRPVEEAEPTAGVAKRAARALRPLMEGSGNNPLLLAALLTGVGVASSL
LIPILGMILVILSRLLTPAOSKMAVIGLPTGVAAAGVYLMGTGEGWQDPAEGHIM
NDATIEHPWVILRTAAVATALYLHMRSRPHG"
5832..6186
/note="6x degenerate repeat:
CCG(T/A)CCG(C/G)GG(A/C)(G/C)"
6293..6955
/gene="SCD12A.06"
6293..6955
/gene="SCD12A.06"
/note="SCD12A.06, possible hydrolase, len: 220 aa; similar
to TR:086530 (EMBL:AL031124) Streptomyces coelicolor
putative hydrolase SCIC2.30. 244 aa; fasta scores: opt:
128 z-score: 155.4 E(): 0.34; 32.9% identity in 167 aa
overlap. Contains Pfam match to entry PF00702 Hydrolase,
haloacid dehalogenase-like hydrolase"
/codon_start=1
/transl_table=11
/product="putative hydrolase"
/protein_id="CAB93397.1"
/db_xref="GI:8249976"
/translation="MASTTRPAPVGFDDMLTIDSRPGIRACVVELSARTGYVDAD

[illegible]

Db 24097 GGGGATGACGCTCTGTCGCCGCGGTCGCCCGCTGCGGACCACTTGCGGAGATCTC 24156
Qy 796 ctgcgtgtgccccttctgtagagcccgccgacccatgctgagcggtgcga 845
Db 24157 CTGCGTACGCGCCCTGTCTTCGCCCGACCGCTACACATGTCGCGTCCCA 24206

RESULT 11
AE003996 12335 bp DNA BCT 06-JUL-2000
LOCUS Xylella fastidiosa, section 142 of 229 of the complete genome.
DEFINITION AE003996 AE003849
ACCESSION AE003996.1 GI:9106790
VERSION
KEYWORDS
SOURCE
ORGANISM Xylella fastidiosa
Xylella fastidiosa
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.

REFERENCE 1 (bases 1 to 12335)
AUTHORS Silvestri, M.L., Siqueira, W.J., de Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M., Tsubako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M., Melands, J., and Setubal, J.C.
TITLE The genome sequence of the plant pathogen Xylella fastidiosa. The Xylella fastidiosa consortium of the Organization for Nucleotide Sequencing and Analysis, Sao Paulo, Brazil
JOURNAL Nature 406 (6792), 151-157 (2000)
MEDLINE 20365717
REFERENCE 2 (bases 1 to 12335)
AUTHORS Simpson, A.J.G., Relnach, F.C., Arruda, P., Abreu, F.A., Acencio, M., Alvarenga, R., Alves, L.M.C., Araya, J.E., Bais, G.S., Baptista, C.S., Barros, M.H., Bonaccorsi, E.D., Bordin, S., Bove, D.M., Briones, M.R.S., Bueno, M.R.P., Camargo, A.A., Camargo, L.E.A., Carraro, D.M., Carreir, H., Colauto, N.B., Colombo, C., Costa, F.F., Costa, M.C.R., Costa-Neto, C.M., Coutinho, L.L., Cristofani, M., Dias-Neto, E., Docena, C., El-Dorriy, H., Facincani, A.P., Ferreira, A.J.S., Ferreira, V.C.A., Ferro, J.A., Fraga, J.S., Franca, S.C., Franco, M.C., Frohme, M., Furian, L.R., Garnier, M., Goldman, G.H., Goldman, M.H.S., Gomes, S.L., Gruber, A., Ho, P.L., Hoheisel, J.D., Junqueira, M.L., Kemper, E.L., Kitajima, J.P., Krieger, E.E., Kuramae, E.E., Laigret, F., Lambais, M.R., Leite, L.C.C., Lemos, E.G.M., Lemos, M.V.F., Lopes, S.A., Lopes, C.R., Machado, J.A., Machado, M.A., Madeira, A.M.B.N., Madeira, H.M.F., Marino, C.L., Marques, M.V., Martins, E.A.L., Martins, E.M.F., Matsukuma, A.Y., Menck, C.F.M., Miracca, E.C., Miyaki, C.Y., Monteiro-Vitorello, C.B., Moon, D.H., Nagai, M.A., Nascimento, A.L.T.O., Netto, L.E.S., Nhani Jr., A., Nobrega, F.G., Nunes, L.R., Oliveira, M.A., de Oliveira, M.C., de Oliveira, R.C., Palmeri, D.A., Paris, A., Peixoto, R.B., Pereira, G.A.G., Pereira Jr., H.A., Pesquero, J.B., Quaggio, R.B., Roberto, P.G., Rodrigues, V., de M. Rosa, A.J., de Rosa Jr., V.E., de Sa, R.G., Santelli, R.V., Sawasaki, H.E., da Silva, A.C.R., da Silva, F.R., da Silva, A.M., Silva Jr., W.A., da Silveira, J.F., Silvestri, M.L.Z., Siqueira, W.J., de Souza, A.A., de Souza, A.P., Terenzi, M.F., Truffi, D., Tsai, S.M., Tsubako, M.H., Vallada, H., Van Sluys, M.A., Verjovski-Almeida, S., Vettore, A.L., Zago, M.A., Zatz, M., Melands, J., and Setubal, J.C.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP 13083-970, Brazil

REMARK FEATURES
Source Location/Qualifiers
1..12335
/organism="Xylella fastidiosa"
/db_xref="taxon:2371"
complement(115..1203)
/gene="XF1723"
complement(115..1203)
/gene="XF1723"
/note="similar to GI1934634 (percent identity: 55 %/query alignment coverage: 89.2 %/subject alignment coverage: 102.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder/Start codon shift: 90"
/codon_start=1

gene CDS
/transl_table=11
/product="sugar-phosphate dehydrogenase"
/protein_id="AAF84532.1"
/db_xref="GI:9106791"
/translation="MGMASTLAAPLIGALSACTARASAAARAIGSGYKHLGRGLKYSRIALGCMNGCELTDENSRFRMSALDSVNFEDTADYVGGGTPEPMKGTSEETIGKRLADKSRDRDTIVATVYOPMGTPNDKTLUSAIHIRACEASLKRKTUHDLYOMHHVDSSTPMQEIWMQAEQVREKLTIVYSSNFAMDIARQGVASRNELGLVSEOSLYNLIORTIELEIVPAVREIGLIPMSPIGMGLGVLAGITTEGRRATPGLOA OIKRPPOLEAVEALCRELGOPAVVALAWVHNHPVTAISGPRVTEQRENLKALS INLSEETLAKIDEIMPGEAPKAYAW"
complement(1437..2426)
/gene="XF1724"
complement(1437..2426)
/note="similar to GI1934634 (percent identity: 55 %/query alignment coverage: 98.2 %/subject alignment coverage: 102.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="sugar-phosphate dehydrogenase"
/protein_id="AAF84533.1"
/db_xref="GI:9106792"
/translation="MNYTHLGRTGLKVSRIALGTMNFGELTDEATSFRTMDTADGIFNEFDADYVGGPQTPDMPKGTSEETVIGNWLAQDSRDKTLVATKYQPEGTGPNKYLASVHINRAACEASLKRKTUHDLYOMHHIDRATPEEIWMQAEQVREKLTIVYSSNFAMDIARQGVASRNELGLVSEOSLYNLIORTIELEIVPAVREIGLIPMSPIGMGLGVLAGITTEGRRATPGLOA OIKRPPOLEAVEALCRELGOPAVVALAWVHNHPVTAISGPRVTEQRENLKALS INLSEETLAKIDEIMPGEAPKAYAW"
complement(2494..2742)
/gene="XF1725"
complement(2494..2742)
/note="hypothetical protein; identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF84534.1"
/db_xref="GI:9106793"
/translation="MPTIIVKIAEGRSOPIDKQELDRLAATMMDVLGSDSSAVSAYVE.DYPMQDMQOYVGPPIETAGEELLRPGGLASPEE"
complement(2758..3525)
/gene="XF1726"
complement(2758..3525)
/note="similar to SPIP50197 (percent identity: 39 %/query alignment coverage: 97.3 %/subject alignment coverage: 99.2 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase"
/protein_id="AAF84535.1"
/db_xref="GI:9106794"
/translation="MNPYDFEKGOVALYTGAAAGMLATARGFAOGAGSVVLADBDGE RAAQEAALIVAGGFAIGLACDSEAOVAALVEKTVAGELDAFNNAGVMAIAPFAEATSEEDRVYIGINLRGVMICRVANVAGCTIDTODIARDVYVGSDEAPRAKLASAPIG ASKHGVILTRAAALEYCKKGRVNAVPGCTIDTODIARDVYVGSDEAPRAKLASAPIG RAGRPETIASAVMLWCSFGSGSTVIGEGLVDSGYVP"
complement(3709..4770)
/gene="XF1727"
complement(3709..4770)
/note="similar to SPIP14941 (percent identity: 42 %/query alignment coverage: 100.0 %/subject alignment coverage: 100.3 %); identified by sequence similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="NADP-alcohol dehydrogenase"

```

gene
/translation="MRCGVYLEAKGTAKKRVKPPVLYGNPAIVLPRTAATCTDVMHI
ATPAALPTKGVYIGHNAVGVYKVEYGVNKKAFPRGGDRVYLPCGMSDRHRPAQGEATY
YQNSRPTSDPSIGGFSEYVRIIDDLALRIIPKVSVDLQVLPDMATPAGTGEV
RDIQIQTETLVMLGDLGVGLMAAAGALRKAGRIIGVDSRPNVLELRAVATIDYF
TGQGTLEKVALTAGQPDVSVLVLAGSGTASETTALKAAPFGGRVACSGFLHDSV
TILEWNVYGMKEKFLTGVDGVODGRDIERLEMLKEGFRDTPRIVSHVLHGWDSLQ
GIELMSRDPAPVIRKPIVYI"
complement(4813. .6156)
/gene="Xf1728"
complement(4813. .6156)
/gene="Xf1728"
/note="similar to GI|290511 (percent identity: 27 %/query
alignment coverage: 84.8 %/subject alignment coverage:
84.0 %): identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="transport protein"
/protein_id="AAE84537.1"
/db_xref="GI:9106796"
/translation="MNPCCARPACGCGSRSRRLFTLPRLSPRAWICROIARGHQO
AVSTLTHRPPIPASSAEDTLPTPALMLAMTGFIAVETTLNAGLIPQIADLATA
PTAGOMVTVYAGSLALAIPLVSLTQTRRKPYMLMALAGFLFNNLTAFSPWPAITL
VAFFLGVSAGSLGMLGLGYSARVYDRIKGRALAVAMWTPIALSLIGVLEFTGTGI
IGMRGAFVLSVVAALVLSAMRMPEVPCQPEPRISISVLSAGSVRSILPTALAM
VTRQMYLTYIAVFAGTILGRADYDILLAFGVAIYGVIMVAGSDVRHLRVYIGSL
VLVAIVLFTSPASALFAPVAPMLMGASFGGANTSIOPTAASDAGEVDYIGAMLT
TWNAGIAAGCALVTLARGASVLPASMLPLIVALLAALRLACRRGRFAGARA"
complement(6153. .7028)
/gene="Xf1729"
complement(6153. .7028)
/gene="Xf1729"
/note="similar to GI|6459673 (percent identity: 61 %/query
alignment coverage: 97.9 %/subject alignment coverage:
87.2 %): identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="phenylacetaldehyde dehydrogenase"
/protein_id="AAE84538.1"
/db_xref="GI:9106797"
/translation="MKIDASTSGOFAIGDLTVNRKGFGMARITGPDWKEPEPHDEA
IRULKRPEIGVDLIDTADSYGPRVSOQLADALHPYGGIKIATKGSVYRPGNSTNP
SMYISQDPAITLKEOCVYMSLKRLEQDIDLMQHLRIDKVPRAEDQFGIRFIDEGILR
HAGLSQYSAIEKARVFPVATVQNTYNNLADRADEDYDLYCEANGIGFIPWPLAG
DLKPGGAVDALAKAKGATAGQIOLAMLLKRSPIVLIPIPTSKVAHEENVAAATIL
SDEEFAIDAAAPRG"
complement(7165. .8055)
/gene="Xf1730"
complement(7165. .8055)
/gene="Xf1730"
/note="similar to SP|P30864 (percent identity: 35 %/query
alignment coverage: 96.6 %/subject alignment coverage:
94.1 %): identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="transcriptional regulator (LysR family)"
/protein_id="AAE84539.1"
/db_xref="GI:9106798"
/translation="MQVYHIGSFMAAAREQNVDPSSVSRVAALAEALGRLPARNTR

```

[illegible]

regulatory protein; secreted beta-galactosidase; serine/threonine protein kinase; tetr-family transcriptional regulator.

SOURCE
ORGANISM
 Streptomyces coelicolor A3(2)
 Streptomyces coelicolor A3(2)
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.

REFERENCE
AUTHORS
 1 (bases 1 to 34074)
 Redenbach, M., Kiser, H.M., Denapate, D., Elchner, A., Cullum, J.,
 Kinash, H. and Hopwood, D.A.
 A set of ordered cosmids and a detailed genetic and physical map
 for the 8 Mb Streptomyces coelicolor A3(2) chromosome
MOL. MICROBIOL. 21 (1), 77-96 (1996)
 97000351

JOURNAL
MEDLINE
 2 (bases 1 to 34074)
 Oliver, K. and Harris, D.
 Unpublished
 3 (bases 1 to 34074)
 Cerdeno, A.M., Parkhill, J., Barrell, B.G. and Randsdram, M.A.
 Direct Submission
 Submitted (23-NOV-1999) Streptomyces coelicolor sequencing project,
 Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
 CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
 David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
 Colney, Norwich, Norfolk NR4 7UH, UK
 On Nov 26, 1999 this sequence version replaced gi:6468236.

COMMENT
 Notes:
 Streptomyces coelicolor sequencing at The Sanger Centre is funded
 by the BBSRC and Beowulf Genomics
 Details of S. coelicolor sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL: <http://www.sanger.ac.uk/projects/S-coelicolor/>)
 CDS are numbered using the following system eg SC7B7.01c. SC (S.
 coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
 strand).
 The more significant matches with motifs in the PROSITE database
 are also included but some of these may be fortuitous.
 The length in codons is given for each CDS.
 Usually the highest scoring match found by fasta -o is given for
 CDS which show significant similarity to other CDS in the database.
 The position of possible ribosome binding site sequences are given
 where these have been used to deduce the initiation codon.
 Gene prediction is based on positional base preference in codons
 using a specially developed Hidden Markov Model (Krogh et al.,
 Nucleic Acids Research, 22(22):4768-4778(1994)) and the Frameplot
 program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the
 correct initiation codon. Where possible we choose an initiation
 codon (atg, gtg, ttg or (att) which is preceded by an upstream
 ribosome binding site sequence (optimally 5-13bp before the
 initiation codon). If this cannot be identified we choose the most
 upstream initiation codon.
 IMPORTANT: This sequence MAY NOT be the entire insert of the
 sequenced clone. It may be shorter because we only sequence
 overlapping sections once, or longer, because we arrange for a
 small overlap between neighbouring submissions.
 Cosmid F81 lies on the Ase-I genomic restriction fragment.
 Location/Qualifiers
 1. 34074
 /organism="Streptomyces coelicolor A3(2)"
 /strain="A3(2)"
 /db_xref="taxon:100226"
 /clone="cosmid F81"
 1. 118
 /gene="SCF81.01"
 /gene="SCF81.01"
 1. 115
 /note="SCF81.01"
 /note="nominal overlap with Streptomyces coelicolor SL5B7"
 <1. 118
 /gene="SCF81.01"
 /note="SCF81.01, ABC transporter (partial CDS), len: >38
 aa: identical to C-terminal region of previously sequenced
 TR:09ZNB0 (EMBL:AB019513) Streptomyces coelicolor ABC

transporter, 584 aa"
 /codon_start=2
 /transl_table=11
 /product="ABC transporter"
 /protein_id="CA61535.2"
 /db_xref="GI:6468413"
 /translation="TVVADAGRVAVGTHRELVTADPLXELATQFLATG"
 complement(127..411)
 /gene="SCF81.02c"
 complement(127..411)
 /gene="SCF81.02c"
 /note="SCF81.02c, unknown, len: 94 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SCF81.02c"
 /protein_id="CA61536.2"
 /db_xref="GI:6468414"
 /translation="MISRRRIYAVVGLAVGTGLAAPMASAAGGPGTCKLSVTNTL
 DSLAVGDMRPAEDRAMPLPSQQLNRRLDLDLQDYTGVSFPGVPALG"
 583..587
 593..958
 /gene="SCF81.03"
 593..958
 /gene="SCF81.03"
 /note="SCF81.03, unknown, len: 121 aa"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein SCF81.03"
 /protein_id="CA61560.1"
 /db_xref="GI:6468262"
 /translation="MGTTAGRGPFDEALRGIEGTAPLISLRARDAEVRVRYDQ
 PSREVLGRDELEMLDDVYSRDLTHMARCVGGDVAFSECTPDGVRVSESM
 LSLRDELVEQTMQAMDE"
 complement(986..1645)
 /gene="SCF81.04c"
 complement(986..1645)
 /gene="SCF81.04c"
 /note="SCF81.04c, possible tetr-family transcriptional
 regulator, len: 219 aa; similar to various Streptomyces
 transcriptional regulators, e.g. TR:09ZBR0 (EMBL:AL035161)
 Streptomyces coelicolor putative transcriptional regulator
 SC967.10c, 272 aa; fasta scores: opt: 189 z-score: 243.3
 E(): 3.5e-06; 29.8% identity in 218 aa overlap. Contains
 match to Pfam entry PF00440 tetr, Bacterial regulatory
 proteins, tetr family"
 /codon_start=1
 /transl_table=11
 /product="putative tetr-family transcriptional regulator"
 /protein_id="CA61559.1"
 /db_xref="GI:6468261"
 /translation="MAGRLRAPTGKRYGKSAQROAERRRFLDAALQLFGDSPCYRA
 TTVAALSEAGLSTROFYEERFTLEDVLAALHLQVNAAEVVRPAAGAGGLPADR
 AAVIFRAYANVATDPDRVRIITFEIVGVSPRLSEORARARARVLDLCAEDRSNAVR
 GEATRDHRLAATATIGVNGVGLHDYSGWVDATLDEVDELVHLLAVLRAPSPSE
 G"
 complement(1421..1564)
 /gene="SCF81.04c"
 /note="Pfam match to entry PF00440 tetr, Bacterial
 regulatory proteins, tetr family, score 18.40, E-value
 0.0023"
 complement(1652..1656)
 1832..1836
 1845..3098
 /gene="SCF81.05"
 1845..3098
 /gene="SCF81.05"
 /note="SCF81.05, possible serine/threonine protein kinase,
 len: 462 aa; low similarity to C-terminus SW:PKWA-THRCU
 (EMBL:U03820) Thermomonospora curvata putative
 serine/threonine-protein kinase PKWA (EC 2.7.1.1), 742 aa;
 fasta scores: opt: 157 z-score: 167.3 E(): 0.06; 23.6%
 identity in 309 aa overlap. Contains possible N-terminal
 signal peptide sequence"

		/codon_start=1	
		/transl_table=11	
		/product="putative serine/threonine protein kinase"	
		/protein_id="CAB6158.1"	
		/db_xref="GI:6468260"	
		[translation="MAPAFSTRHLGVAALATLTVASPAATAAAPADPAGILREYMEVG NMNMGTDVLEISTGDFARIGRIDVPDPKARMAEINADPIKMIYFQAIRNGVEGHQO LVDMYSSTPDTSVSVSPSPADVASVDLTGDNEMEPGSGVSDHAAVSPOSTRA VSAASTNHVLDIIRTEGOVSGFGTGPKPHENITRTGRGYIMNSIGDWTALDAPL DMTGDRRITYVDAKTRYOKAYVIDMRRLDALIGDSADVAPRAVEFPEDSKILFOVS DFVFNEFEIDVAGDKTRITRKTLPKNPATSEDTTHWNDRSHGMSWDGARICVAGTM DVAVANVRBDILEEGPLVPAPKPWAIWVSDDGTCHVYESGADQVNTAIDEATGERKS VPMDHEORVRIIGVIAADWTPGSAG" complement(3339) .3839) /gene="SCF81.06c" complement(3339) .3839) /note="SCF81.06c, hypothetical protein, len: 166 aa; similar to various hypothetical proteins, e.g. TR.AAB36153 (EMBL:AEO01768) Thermotoga maritima conserved hypothetical protein TM1076, 157 aa; fasta scores: opt: 293 z-score: 385.9 E(): 4e-14; 39.1% identity in 133 aa overlap"]	
CDS		/codon_start=1	
		/transl_table=11	
		/product="hypothetical protein SCF81.06c"	
		/protein_id="CAB6157.1"	
		/db_xref="GI:6468259"	
		[translation="MRLILTSDFHLPRAKELPAPLAETIPRADVLHAGDWDTATL DLSSRRLGVGNNDGPALRARLPDEVAVADLGFRGVHETGAOGREARCAAR EPGLDILVEGHSHPMTTATATGLRLNLNPSPTDRNRQPPRTYTATVADGRLDVEL HRLEPR" 3962.4543 /gene="SCF81.07" 3962.4543 /gene="SCF81.07" /note="SCF81.07, possible hydrolase, len: 193 aa; similar to TR:P94573 (EMBL:P82987) Bacillus subtilis hypothetical 21.1KD protein, 189 aa; fasta scores: opt: 504 z-score: 582.8 E(): 4.4e-25; 48.4% identity in 190 aa overlap and to SW:EHTB_ECOLI (EMBL:M24148) Escherichia coli isochorismatase (EC 3.3.2.1) (2,3 dihydroxy-2,3 dihydroxypentanoate synthase) EntB, 285 aa; fasta scores: opt: 205 z-score: 242.1 E(): 4.1e-06; 26.8% identity in 183 aa overlap. Contains match to Pfam entry PF00857 Isochorismatase, Isochorismatase family" /codon_start=1 /transl_table=11	
Query Match	21.3%	Score 213.2;	DB 3; Length 34074;
Best Local Similarity	55.3%	Pred. No. 2.3e-15;	
Matches 488;	Conservative 0;	Mismatches 373;	Indels 21; Gaps 3;
OY	26	ccgcgagcgagccaccgcgggggaagaatctctgcacgttcggaacagcgctgctgcacagg	85
Db	33157	CCGGCGATCGGGGTGGGGCGCGTAGAGGCCGCACCTCCTCGATCATCCGTCGCTCAGC	33216
OY	86	gtcatcgcggagacgacttaagcgcgagagcgagagctctcgcgggggttcgcggcgcaatacgy	145
Db	33217	GTTGACTTGGACCCTCCGCCACCGCGCTCTCTCAAGTGGCCCGCCGGAGAGCCGCACAGATC	33276
OY	146	gcgccgcgatgcccggcgcgagacaagcacccattgcagaccccaacctgcggcgagtctctcg	205
Db	33277	GGCGCGGTACCGTCTCTGTCGTGTCAGCACAGGCCAACAGGCGCACCCCGGCGCGCGACG	33336
OY	206	cggaggttgcgcaaaacttctcgttagcgtctcgtatcgcgcggcgcgaggaagcgaaacagc	265
Db	33337	CCGGCTCCCCCGCGATCGGGGTGACGGCTCTCACCAAGGCGCGGCTCCCTCTCGGTAC	33396
OY	266	aactcgcagacgccccctcgcgcgaattacaacgcgcggttcgcccgcgacagttctcacaagct	325
Db	33397	AGGGTGTGTCGCCAAGGTGTGTGGT-----GGCTGTGCGCCGGGTGGGGTGTGCCAGTGG	33450
OY	326	ccgctgagcaagcgcgctgcaagcgcgaccaaagcagaaagacgcgagaccgctaagcgtgc	385

	Db	33451	CGGGTGAAGGCGGCCCCCGGCGAGGCGGACTCCACAGGGGCAGACGCCGCAAGCCTTGATCCGG	33510
OY	386	gcgagcgagcagcaactccacgctcggcggtgcggagccgccagttgttaacagcattgttg	445	
Db	33511	CAGAGGGGGAGCATMTCGGGCTCTCTCCGCGCGGTAGACACAGGTTGTAGTGGTTCTGCATG	33570	
OY	446	gagagcatgcccagggaagtgcggcggggggagtccttctctgcgcggcgagatggtgcag	505	
Db	33571	GACACGAACTTGTCACAGCCGTCGCCGCTCGGGGGTGTACTGCATCTTGAGAACTGCCAG	33630	
OY	506	ccccgaaattcgaacgagcgacgtagagagaacctgcgtgcgcgagagagctgtccatg	565	
Db	33631	GGGTACATCAGAACTGGCCCCCGATTGATAGCGGACCTTAGCCGCTTAGCACGATGCAGAC	33690	
OY	566	gcttcgcaactgctccacagcgcggaaccggtgatagtgtgatctggtgaagatctg	625	
Db	33691	GGCTCATCTGTTCTCTCGACCCGGGTGTCCGGGTGGAAGCGGTGATCTGGTAGAGTTCG	33750	
OY	626	atgttgtcaaacccacagcctgcgcagcgacatccctcgcagagagagcatgtatgtccgcgc	685	
Db	33751	ACGTAGTGGCGTCCCAGGCGCGCCAGAGCGCTGATGTCGATCCTGGTATCACCACCCCTTGGCG	33810	
OY	686	gacagcccgctgtgtttaagcgcgtcgcgtatcttcgtccgcgcgacctgtgtgcgagcag	745	
Db	33811	GAGAGTCCGGCCCGCTTTGGGCCCGGGGGCGCAATCCGGCGCTTACCTTGGTCCGAGGACAG	33870	
OY	746	gtgtctctgcgcgcgtccgcgcgcctgcggagcgacgaacctgtccacagctcctcgtgtg	805	
Db	33871	ATGTGTCGGCGGCG---GGCGAATTCGGCGCAATGCTTGGCCGACATCTCTCTGGTGGTG	33927	
OY	806	ccctctgtagagccgcgcgcgcgttacatgtctgcgcgtgtgtgaggaatttatgccgcggtcc	865	
Db	33928	CCGTGCGA-----GTACACGTTGGGGGTGCMAAGAAGTGAAGCCGGCGCTTC	33975	
OY	866	cggcgctgtcatcatcagagcgccgcgtctgtctcctgcag	907	
Db	33976	AAGCGCTGCGGATCAGCGGACGCGAGCGCGGCTGTGTGACAG	34017	
RESULT	13			
SCKL3				
LOCUS	SCKL3	39896 bp	DNA	BCT
DEFINITION	Streptomyces coelicolor cosmid K13.			26-FEB-2001
ACCESSION	AL451182			
VERSION	AL451182.2	GI:13162229		
KEYWORDS	ABC transporter ATP-binding protein; aldoketoreductase; amino acid deaminase; deaminase; dehydrogenase; DNA-binding protein; integral membrane transport protein; marR-family transcriptional regulator; membrane protein; metB, cystathionine gamma-synthase; narG3, nitrate reductase alpha chain; narH3, nitrate reductase beta chain; narJ3, nitrate reductase gamma chain; narJ3, nitrate reductase delta chain; oxidoreductase; peptide methionine sulfoxide reductase; regulatory protein; tetR-family transcriptional regulator; transcription elongation factor.			
SOURCE	Streptomyces coelicolor			
ORGANISM	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
REFERENCE	1 (bases 1 to 39896)			
AUTHORS	Redenbach,M., Kleiser,H.M., Denapate,D., Eichner,A., Cullum,J., Kinash,H. and Hopwood,D.A.			
TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome			
JOURNAL	Mol. Microbiol.	21 (1),	77-96	(1996)
MEDLINE	97000351			
REFERENCE	2 (bases 1 to 39896)			
AUTHORS	Oliver,K. and Harris,D.			
JOURNAL	Unpublished			
REFERENCE	3 (bases 1 to 39896)			
AUTHORS	Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.			
TITLE	Direct Submission			
JOURNAL	Submitted (11-Dec-2000) Streptomyces coelicolor sequencing project.			

ALRELARYRSEGVSLALPCVATGVGLDMSEVKPLVVRHLGDLLEPVILKEYVRKG
VAAEKLIA"
BASE COUNT 1741 a 3254 c 3569 g 1750 t
ORIGIN

Query Match 18.5%; Score 185.2; DB 1; Length 10314;
Best Local Similarity 52.8%; Pred. No. 4,4e-12;
Matches 459; Conservative 0; Mismatches 398; Indels 15; Gaps 2;

```

34  ggcacccggggggaagatctctcagttcgagacagcgccgagctccagagtcacgc 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
550  ggcggcgccgctgacgacgcccctccacagcagcagcagcagcagcagcagcagc 581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94  ggcagcctcagcgagagtcagagctcaggggttcagcgagcagcagcagcagcagc 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5890  CAGCGCGCGGAGGAGATCTTCAGAGTGATGGCGCTGACAGCGCGGAGATCGCGCGC 5831
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154  gatcgcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5830  GACACCGCTCCGCGCAAGCAGCAGCGCAGCGAGTTGCGCGGTGGAAGCCGCGGCG 5771
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214  gcgcagaaactctcagagcctcagcagcagcagcagcagcagcagcagcagcagc 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5770  AGCGGCGAGCTTCCCAACGCGCGGAGGTCGACTCGACTCTCCCTCGCGCGGTACCA 5711
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274  acggcctcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 333
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5710  GCGCGGGGCGTTCGTCGTCGCGGTCGAGAGTCCGCGCGCTTCGCGCTCGCGCGAG 5651
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
334  cagcgccgctgcagcgagcagcagcagcagcagcagcagcagcagcagcagcagc 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5650  CAGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 5591
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
394  cagcagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 453
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5590  CACGAGGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 5531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
454  gccacggagatgagcgagcgagcgagcagcagcagcagcagcagcagcagcagc 513
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5530  TTGGCGCAGCGCGCGGTGGCGCGCTCGAAGCGAGCATCTTCATGAAGCGCGCGCAT 5471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
514  gttcagagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 573
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5470  CGAGGAGCGCGCGGAGTACGAGCAGCTTGCGCACACGACGAGGTGTCGAGGTTTGC 5411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
574  caactctccacagcgagcgagcgagcagcagcagcagcagcagcagcagcagc 633
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5410  GCTTCTCATCGGGGTGTCGAGTGCAGACGATGCATCGAAGAGATCCAGGTAGTC 5351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
634  gacgccagcctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 693
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5350  CGTACCAACCGCGCGAGCGAGCATCCAGCGCTGCAAGCGAGGTGCGCGCACAGGCC 5291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
694  gctgtgttgcagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 753
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5290  GCAAGCGCTTGGCGCTGTCGTCATGGAAGAACGCGCTTGTCGATGACCAAGCTCTC 5231
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
754  gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5330  GCGCTGCGCGAAGTCCCTC---AGCGCACGCCCGAGATCCGCTGCTGCGCGGTGA 5174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
814  gacgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 873
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5173  -----ATAGATATCGGCACCTGGAAGAAAGTCATCCCGGCTTCACAGCGCTC 5126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
874  gtccatcagcgagcgagcgagcagcagcagcagcagcagcagcagcagcagcagc 905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
5125  GCGAATGAGGAGCGCGCGCGCTTCCTGTCGA 5094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15
AP03003/c

LOCUS AP03003 349116 bp DNA BCT 03-FEB-2001
DEFINITION Mesorhizobium loti DNA, complete genome, section 10/21, complete
sequence.
ACCESSION AP03003 BA000012
VERSION AP03003.1 GI:11994978
KEYWORDS HTG.
SOURCE Mesorhizobium loti (strain:MAFF303099) DNA.
ORGANISM Mesorhizobium loti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.
REFERENCE
1 (sites)
Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S.,
Watanabe, A., Iidesawa, K., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A.,
Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M.,
Takeuchi, C., Yamada, M. and Tabata, S.
Complete genome structure of the nitrogen-fixing symbiotic
bacterium Mesorhizobium loti
DNA Res. 7 (6), 331-338 (2000)
JOURNAL
MEDLINE
21082930
REFERENCE
2 (bases 1 to 349116)
AUTHORS
Kaneko, T.
TITLE
Direct Submission
JOURNAL
Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp,
URL: http://www.kazusa.or.jp/rhizobase/, Tel: 81-438-52-3935,
Fax: 81-438-52-3934)
FEATURES
Location/Qualifiers
source
1..349116
/organism="Mesorhizobium loti"
/strain="MAFF303099"
/db_xref="taxon:381"
BASE COUNT 63710 a 112865 c 109235 g 63306 t
ORIGIN

Query Match 18.1%; Score 180.8; DB 2; Length 349116;
Best Local Similarity 52.8%; Pred. No. 1.4e-12;
Matches 471; Conservative 0; Mismatches 402; Indels 19; Gaps 3;

```

34  ggcacccggggggaagatctctcagttcgagacagcgccgagctccagagtcacgc 93
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151381  GGAGGCGAGCGCGGTGTCGAGTGCAGACCTGCGCAGATGCTCCGCTGAGCTTACTCC 151322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94  ggcagcctcagcgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151321  CTTGCGCGCGAGGTTCTGTTGAGACTGTCATGCTGAGCTGAGCGCGCGGATGACCTT 151262
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
154  gatcgcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151261  CAGC-AMGCGTGGCGAGCATATAGCGAGCGCGCAGCGTGGCAAGCTGACGTGATCTT 151203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214  gcgcagaaactctcagagcctcagatcgccgagcgagcagcagcagcagcagcagc 273
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151202  CTTGGCATCTCGCGCATGACGCGGAGCGCGCGCAGCGCGGCTGTTGACCGAGG 151143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
274  acggcctcgcgcg-----acttcacgcygagcgcgcgcgcgcgagcagcagcagc 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151142  GAAATTGAAGAGCACGACGCGCGCTTCGCAATGCGCGCGCGCGCGAGACCTTATTC 151083
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
328  gctgagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 387
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151082  GGAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151023
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
388  ggcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 447
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
151022  GAGCGGACAGCATCTCGCGCTCGAGAGTGGCTCGCGCGCATGGAATATAGACGTGATGT 150963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
448  gaccatgcccagagagtgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcg 507
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
150962  CTGGAAGCGGCGGAGAGCCCTTGGGCTGGCGATGCCGAGCGCGCTTGGCGATGGCCATG 150903
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2001, 09:23:55 ; Search time 28.5 Seconds

(Without alignments) updates/sec
1545.880 Million cell

Title: US-09-463-705A-2

Perfect score: 1728

Sequence: 1 MTTDATATVRLGRSALTSTR.....LDEIFPAVSGAAPPAAWLO 333

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_16:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_prodent:*
13: sp_unclassified:*
14: sp_vertebrate:*
15: sp_virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1728	100.0	333	2	033936
2	1415.5	81.9	330	2	09F833
3	1286.5	74.5	329	2	09XC70
4	1056	61.1	347	2	09S0M9
5	1025	59.3	323	2	09F825
6	841	48.7	330	2	09F255
7	804.5	46.6	329	2	09PC01
8	791.5	45.8	362	2	09PC05
9	763	44.2	316	2	09PC05
10	745	43.1	354	2	09KX18
11	495	28.6	354	2	044328
12	472.5	27.3	323	2	09I228
13	428.5	24.8	336	2	09I228
14	421	24.4	307	2	09RUC4
15	414.5	24.0	352	2	09K057
16	408	23.6	316	2	09HXR2
17	399	23.1	324	1	09HPH9
18	393	22.7	344	3	059826
19	387	22.4	343	1	09H190

20	384	22.2	315	2	09RS10	09rs10 delinococcus
21	380.5	22.0	328	10	023016	023016 arabidopsis
22	379.5	22.0	328	10	049847	049847 egeria dens
23	376.5	21.8	328	10	039151	039151 arabidopsis
24	373	21.6	345	2	09HXY1	09hxy1 pseudomonas
25	371.5	21.5	342	3	002895	002895 saccharomyces
26	363	21.0	419	4	014722	014722 homo sapien
27	363	21.0	419	6	09XT31	09xt31 oryctolagus
28	360.5	20.9	330	10	082064	082064 solanum tub
29	359.5	20.8	408	6	028528	028528 mustela put
30	356	20.6	401	11	063277	063277 mus musculu
31	355	20.5	327	2	09RS66	09rs66 delinococcus
32	344	19.9	401	13	09PRL1	09prl1 gallus gall
33	340	19.7	319	2	09WYR5	09wyer5 thermotoga
34	339.5	19.6	336	1	034185	034185 halobacteri
35	338.5	19.6	351	3	09P702	09p702 schizosacch
36	338	19.6	336	1	09HOU7	09hou7 halobacteri
37	337.5	19.5	387	2	09PGD5	09pgd5 xyella fas
38	333	19.3	453	10	p93818	p93818 arabidopsis
39	331	19.2	367	13	09PTM5	09ptm5 xenopus lae
40	330	19.1	367	6	09NOV9	09nov9 oryctolagus
41	329.5	19.1	330	2	09EMF0	09emf0 streptomyce
42	329	19.0	344	2	09X5G7	09x5g7 streptomyce
43	328	19.0	367	6	027955	027955 bos taurus
44	328	19.0	367	11	064284	064284 mus musculu
45	325.5	18.8	348	2	09EMJ0	09emj0 streptomyce

ALIGNMENTS

RESULT 1
033936 PRELIMINARY: PRT; 333 AA.
ID 033936;
AC 033936;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE ERBBI1.
GN ERBBI1.
OS Saccharopolyspora erythraea (Streptomyces erythraeus).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Pseudonocardiales; Pseudonocardaceae;
OC Saccharopolyspora.
OX NCBI_TaxID=1836;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL2338;
RX MEDLINE=98015410; PubMed=9353926;
RA Summers R.G., Donadio S., Staver M.J., Wendt-Pienkowski E.,
Hutchinson C.R., Katz L.;
RT "Sequencing and mutagenesis of genes from the erythromycin
biosynthetic gene cluster of Saccharopolyspora erythraea that are
involved in L-mycarose and D-desosamine production.";
RL Microbiology 143:0-0(0).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL2338;
RA Salah-Bey K., Doumith M., Michel J.M., Haydock S., Cortes J.,
Leadlay P.E., Raynal M.C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; U77454; AAB84068.1; -;
DR EMBL; Y14332; CAAT4709.1; -;
SQ SEQUENCE 333 AA; 35867 MW; 6F912355F21B6F8F CRC64;

Query Match 100.0%; Score 1728; DB 2; Length 333;
Best Local Similarity 100.0%; Pred. No. 3.1e-114;
Matches 333; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MTTDATATVRLGRSALTSTRMLGTVNFGSRVEDDALARLMDHARDGNCIDTADMGW 60
DB 1 MTTDATATVRLGRSALTSTRMLGTVNFGSRVEDDALARLMDHARDGNCIDTADMGW 60

QY 61 RLYKGTTELVRWLAAGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRL 120
 DB 61 RLYKGTTELVRWLAAGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRL 120
 QY 121 GVDHIDVYQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIAAOGENARRHS 180
 DB 121 GVDHIDVYQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIAAOGENARRHS 180
 QY 181 LGWVSHOCLNLAVRHALEVLPAQAQYGLGVFAMSPPLHGGLLSGALEKLAAGTAVKSAQ 240
 DB 181 LGWVSHOCLNLAVRHALEVLPAQAQYGLGVFAMSPPLHGGLLSGALEKLAAGTAVKSAQ 240
 QY 241 GRAOVLPLSRPAIEAVEKFCRNLDGEDPAEVGLAWLVSRRGIGAGVIGPTPTPQDLSALK 300
 DB 241 GRAOVLPLSRPAIEAVEKFCRNLDGEDPAEVGLAWLVSRRGIGAGVIGPTPTPQDLSALK 300
 QY 301 ASAMTDEQALSELDETFPAVASGGAPEAWLQ 333
 DB 301 ASAMTDEQALSELDETFPAVASGGAPEAWLQ 333

RESULT 2

QY 09F833 PRELIMINARY: PRT: 330 AA.
 ID 09F833;
 AC 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-MAR-2001 (Tremblrel. 16, last annotation update)
 DE TDP-4-KETO-6-DEOXYHEXOSE 2,3-REDUCTASE.
 GN MEGDVII.
 OS Micromonospora megalomicea subsp. nigra.
 OC Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Micromonosporaceae;
 OC Actinomycetales; Micromonosporaceae; Micromonosporaceae;
 OC Micromonospora.
 NCBI_Taxid=136926;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NRRL3275;
 RX MEDLINE=20430101; PubMed=10972798;
 RA Volchegursky Y., Hu Z., Katz L., McDaniel R.;
 RT "Biosynthesis of the anti-parasitic agent megalomicin: transformation
 of erythromycin to megalomicin in Saccharopolyspora erythraea.";
 RL Mol. Microbiol. 37:752-762(2000).
 DR EMBL: AE263245; AAG13914.1;
 SQ SEQUENCE 330 AA; 35434 MW; 32DA7C2DDEF60997 CRC64;

Query Match 81.9%; Score 1415.5; DB 2; Length 330;
 Best Local Similarity 81.4%; Pred. No. 2.2e-100;

Matches 267; Conservative 26; Mismatches 34; Indels 1; Gaps 1;

QY 2 TTGA-ATHVLRGSAITSLRLMTGVNFGSRVEDDQALRLMDHARDCINCLDTADMYGW 60
 DB 3 TTGA-ATHVLRGSAITSLRLMTGVNFGSRVEDDQALRLMDHARDCINCLDTADMYGW 60
 QY 61 RLYKGTTELVRWLAAGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRL 120
 DB 61 RLYKGTTELVRWLAAGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRL 120
 QY 121 GVDHIDVYQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIAAOGENARRHS 180
 DB 121 GVDHIDVYQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIAAOGENARRHS 180
 QY 181 LGWVSHOCLNLAVRHALEVLPAQAQYGLGVFAMSPPLHGGLLSGALEKLAAGTAVKSAQ 240
 DB 181 LGWVSHOCLNLAVRHALEVLPAQAQYGLGVFAMSPPLHGGLLSGALEKLAAGTAVKSAQ 240
 QY 241 GRAOVLPLSRPAIEAVEKFCRNLDGEDPAEVGLAWLVSRRGIGAGVIGPTPTPQDLSALK 300
 DB 241 GRAOVLPLSRPAIEAVEKFCRNLDGEDPAEVGLAWLVSRRGIGAGVIGPTPTPQDLSALK 300
 QY 301 ASAMTDEQALSELDETFPAVASGGAPEAWLQ 333

DB 303 AAEITLGEELRELEAFPPAPADGVPV 330

RESULT 3

QY 09XC70 PRELIMINARY: PRT: 329 AA.
 ID 09XC70;
 AC 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, last annotation update)
 DE NDP-HEXOSE 2,3-ENOYL REDUCTASE TYLCII.
 GN TYLCII.
 OS Streptomyces fradiae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Streptomyces.
 OC Actinomycetales; Streptomyces; Streptomyces.
 NCBI_Taxid=1906;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-T59235;
 RA Bate N., Butler A.R., Smith I.P., Cundliffe E.;
 RT "The mycarose-biosynthetic genes of Streptomyces fradiae, producer of
 tylosin.";
 RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF147704; AAD41821.1;
 DR HSSP: P14550; 2ALR.
 DR InterPro: IPR001395;
 DR Pfam: PF00248; also ket.red. 1.
 DR PRINTS: PR00069; ALDKETREDTASE.
 SQ SEQUENCE 329 AA; 36144 MW; 14BIE58076AEC09B CRC64;

Query Match 74.5%; Score 1286.5; DB 2; Length 329;
 Best Local Similarity 74.7%; Pred. No. 1.6e-90;

Matches 242; Conservative 31; Mismatches 50; Indels 1; Gaps 1;

QY 8 HVLGRSALITSLRLMTGVNFGSRVEDDQALRLMDHARDCINCLDTADMYGW 67
 DB 5 YVIGGATLVSRMLGTGVNFGSRVEDDQALRLMDHARDCINCLDTADMYGW 64
 QY 68 EELVGRWLAAGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRLGVPHIDV 127
 DB 65 EELVGRWLAAGGRRDVTATKVGEMSERVNDGSLARHIIASCEGSLRLGVPHIDV 123
 QY 128 YQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIAAOGENARRHS 187
 DB 124 YQMHIDRSAPMDEWQAMDSLVASGVSYGSSNFRAGWHIAAOGENARRHS 183
 QY 188 CLYNLAVRHALEVLPAQAQYGLGVFAMSPPLHGGLLSGALEKLAAGTAVKSAQ 247
 DB 184 CLYNLAVRHALEVLPAQAQYGLGVFAMSPPLHGGLLSGALEKLAAGTAVKSAQ 243
 QY 248 PSIRPAIEAVEKFCRNLDGEDPAEVGLAWLVSRRGIGAGVIGPTPTPQDLSALK 307
 DB 244 PSIRPAIEAVEKFCRNLDGEDPAEVGLAWLVSRRGIGAGVIGPTPTPQDLSALK 303
 QY 308 EQALSELDETFPAVASGGAPEAWLQ 331
 DB 304 EEBHRELEALFPVVGSGGEVPEAW 327

RESULT 4

QY 09S0N9 PRELIMINARY: PRT: 347 AA.
 ID 09S0N9;
 AC 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, last annotation update)
 DE TDP-4-KETO-6-DEOXY-L-HEXOSE 2,3-REDUCTASE.
 GN AVEBIII.
 OS Streptomyces avermitilis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinomycetales; Streptomyces.
 OC Actinomycetales; Streptomyces; Streptomyces.


```

Query Match      45.8%; Score 791.5; DB 2; Length 362;
Best Local Similarity 48.6%; Pred. No. 1,2e-52;
Matches 162; Conservative 46; Mismatches 112; Indels 11; Gaps 4;

OY 5 AATHVRLGRSALLTSRLMGTGTFNSGVRDEDDALRLMDHARDRCINCLDTADMYGWM----60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 33 SGAKFKHGRGTGLKYSRILGCMGFGELTDEPNFRSMSEALDSGVNLEPDTADYVGGPQT 92
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 61 RLYRGH--TEELVGRWLAAGSGREDIVYLAATKVGEMSERVNSGLSRHIIASCEGSLR 118
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 93 DMERGFTSEETIGRWLAQDRSRDRITATKRYQPMGTGNDYLLSAVHTRRACEASTLK 152
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 119 RLGVHDIVYQMHIDISAPDEWQAMDSLVAAGKSYGSSNPAGHIIAAGENARR 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 RLKTDHIDLQMHVDSSTPQOELTQWAMEQLVREKRTTYGSSNFAAMDIAAGVAESR 212
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 179 HSLGMSHOCILYNLAVRAELEVLPAAQAVGLGVFANSPHLGHLISGALTEKLAAGTAVKS 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 213 NPLGLVSEQSLYNLIQRTIELEVIPAVRELIGLIPMSPIGMGLGVLGKITEG---RR 269
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 239 AQAQAVLPSLRPAIEAYEKEFCNLGEDPAEVLAWLSPRTAGAVIGRPTPEQLDSA 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 270 ATPGLQAOIQKFRPOLAEYEAICRELQCPAPVALAWVHNPTTAISGPRVEQOMREN 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 299 LKASAMTLDQALSELDEIFPAVASGGAPEAW 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 330 LKALSLMLSETLAKLDEIWP--GPGGEAPRAY 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
005408 PRELIMINARY; PRT; 316 AA.
AC 005408;
DT 01-JUL-1997 (TEMBLrel. 04, Created)
DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)
DT 01-NOV-1998 (TEMBLrel. 08, Last annotation update)
DE YRPG.
GN YRPG.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,
RA Duesterhoeft A., Ehrlich S.D.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
MDLINE=98044033; Pubmed-9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Brann M., Bridnell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.O., Daniel R.A.,
RA Danilozot F., Devaine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entlin K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Filiz S.Y., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Gilm S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Huilo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetler P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nkai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwolik S., Prescott A.M.,
RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadie Y.,
RA Sato T., Scanlan E., Schleich S., Schreuter R., Scoffone F.,

```

```

RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takematu K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "the complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: 299117; CAB14626.1; -
SQ SEQUENCE 316 AA; 35475 MW; 207C59D9B0A9CED1 CRC64;

Query Match      44.28%; Score 763; DB 2; Length 316;
Best Local Similarity 50.6%; Pred. No. 1,4e-50;
Matches 161; Conservative 45; Mismatches 100; Indels 12; Gaps 6;

OY 19 SRLMLGTGTFNSGVRDEDDALRLMDHARDRCINCLDTADMYGWRLYKGTTELVGRWLAAG 78
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4 SRLGLTGMNFVDVDEKFAFRIMDEALDNGIQFDTAIYGMGNAGTIESIIKWFAG 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 79 GGRREDIVYLAATKVGEMSERVNS----GLSRHIIASCEGSLRRLGVHDIVYQMHII 133
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 GQREKAVYLAATKVEYEPISD-PNDGPNDRKGLSLKIRRHLEGSIAKRIOTDIHELQMHII 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 134 DRAPWDEWQAMDSLVAAGKSYGSSNFAGMHIAAQAENARHSLGMSHOCILYNLA 193
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 DRPPPEIWEAFETQVRSKGVYDYGSSNFAGMHLYKRAEAKRRFELVTEQHKYSL 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 194 VRHALEVLPAQAQAVGLGVFANSPHLGHLISGALTEKLAAGTAVNSAGRAOVLPLSLRA 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 ERTAEMEVLPAARDLGLGVANSPLAGLGLGKALKSNAQT--RTAK-RAD-LIEKHRLQ 238
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 254 IEAVEKFCRNIGEPDAEVLAWLSPRTAGAVIGRPTPEQLDSAMTLDQALSE 313
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 LEKTSDLCKELGEKANVALAWLANVLTAPITIGPRVDELRTITAVELSDKEILRM 298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 314 LDEIFPAVASGGAPEAW 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 299 LNDIFP--GPGGEPEAY 314
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
09KXLB PRELIMINARY; PRT; 284 AA.
AC 09KXLB;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE POTATIVE REDUCTASE.
GN SCD12A.22C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]

```


DB	232	ACTDEGRLTYDEDOAAERVREIAGRGLSPAOVALMVRNPVSPITGVTKPAOL	291
QY	296	DSALKASAMTLEDOALSELDEIF-----PAVAS 323	
DB	292	AAIAAAYDVGLEDEAAYLEEPPYQPHAEAYLEESFYKARPAAS 335	
RESULT	14		
Q9RJCA			
ID	09RJCA	PRELIMINARY;	PRT; 307 AA.
AC	09RJCA		
DT	01-MAY-2000	(TREMBLrel. 13, Created)	
DT	01-MAY-2000	(TREMBLrel. 13, last sequence update)	
DT	01-JUN-2000	(TREMBLrel. 14, last annotation update)	
DE	PURIFIED ALDO/KETO REDUCTASE (FRAGMENT).		
GN	SCF81.28C.		
OS	Streptomyces coelicolor.		
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.		
OC	Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.		
OX	NCBI_TaxID=1902;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RA	Oliver K., Harris D.;		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RA	Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-A3(2);		
RX	MEDLINE=97000351; Pubmed=8843436;		
RA	Redenbach M., Kiser H.M., Denapate D., Elchner A., Cullum J.,		
RA	Klinasht H., Hopwood D.A.;		
RT	"A set of ordered cosmids and a detailed genetic and physical map for		
RT	the 8 Mb Streptomyces coelicolor A3(2) chromosome.";		
RL	Mol. Microbiol. 21:77-96(1996).		
DR	EMBL: AL133171; CAB61562.2; .		
DR	InterPro: IPR001395; .		
DR	InterPro: IPR001822; .		
DR	Pfam: PF00248; aldo_ket_red; 1.		
DR	PRINTS: PR00069; ALDKETREDTASE.		
DR	PROSITE: PS00398; RECOMBINASE_2; UNKNOWN_1.		
FT	NON TER		
SO	SEQUENCE 307 AA; 34107 MW; EC2D2594595FBI8 CRC64;		
	Query Match	24.4%; Score 421; DB 2; Length 307;	
	Best Local Similarity	36.0%; Pred. No. 1,8e-24;	
	Matches 102; Conservative 46; Mismatches 127; Indels 8; Gaps		
QY	34	DDDALR-LMDHARDGICLDTADMYGWLKYGHTBELVGRMLAOGGGRREDTVLATKVG	92
DB	21	DEAARPLRLQALEAGVTFEDTANYVS---DGTSEELVKALAD-FARRDDIVLATKVN	75
QY	93	GMSERVNDGSLSAHHIIASCGSILRLGVHDIIDYGMHIDRSAPMDVWQAMDSIVAS	152
DB	76	GMRKPGPNAGISRAAVMTEDHSIRRLGTDVLDYLQHRFPDPHPPVEETMEALHDLVKA	135
QY	153	GVSYVYSSNFGWIIAAQENNAARHSIGWVSHOCLYLAVRHALEVLPPAAQVYGLCV	212
DB	136	GKVRIGAGSSMTAMQPSKMQYTAERHGWTKFYSMNHNILVYREERERLPLICADQGVGV	195
QY	213	FAMSEPLHGLGSGALEKTLAAGTAVKSAOGRQVLLPSRLPAIEAYEKFCRLGEDPAEYV	272
DB	196	LPMSPLARQLRDMW-DATGRSATDTFG-SPLYQEGDRAVVEAVTIRLAGEGVPRARVA	253
QY	273	LAWVLSRPGIAGAVIGRPPEQDLSALASAMTLEDOALSELD	315
DB	254	LAMLLHODIVTAPRIGASRPHGLDEAVAVELTLETELEEL	296

```
RESULT 15
09KU57 PRELIMINARY: PRT: 352 AA.
ID 09KU57
AC 09KU57
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE OXIDOREDUCTASE TAS, ALDO/KETO REDUCTASE FAMILY.
GN VC0667.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406633; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL: AE004153; AAF93832.1; -
DR TIGR: VC0667; -
DR InterPro: IPR001395; -
DR Pfam: PF00248; aldo_ket_red: 1.
SQ SEQUENCE 352 AA; 39549 MW; 9E37E3F477BAEA98 CRC64;
```

Query Match 24.0%; Score 414.5; DB 2; Length 352;
Best Local Similarity 34.4%; Pred. No. 6.8e-24;
Matches 116; Conservative 56; Mismatches 140; Indels 25; Gaps 7;

```
OY 5 AATHVRLGRSALLTSLRMVTFNFGSRVEDDALLRLMDHARDGINCLDTADMY---GWR 61
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 8 AMQYTKLPHSSLEISKICIGTWTFFGQNSQADAFQOLDVALERGYNFTDAEMYPPTA 67
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 62 LYKGTTELVGFWLAOGGGRREDTVLATKVGEMS-ERYNDS-GLSARHIIASCEGSLRR 119
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 68 OTGKTEEFITGMWLAK-SGKREKIYATKAGPRNVPYTRDKMALDHRNIHQAVDDSLR 126
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 120 LGVDHIDVYQMHIDRSA-----PMD-----EVMQAMDSLVASGKSYVGSN 162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 127 LQTDYIDLYQLHWPQRTNFGQLNVPYPPDKQEEVTLLETLALNDLVRMGKRYIGVSN 186
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 163 FAGMHIIAAQENAAARRHSIGMVSQCCLVNLAVRHALEVLPAQAQVGLGFVAFMSPLHGGL 222
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 167 ETPMGVMSYLRALAEKHELPRIVSIONPYLLNRSFEVGLAEISHLEGVKLAYSPLAFGA 246
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 223 LSGALEKLA--AGTAVKSAQGAOVLBLSRPAIEAYEKFCRNIGEDPAEVLAWVLSRP 280
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 247 LSGKYLNGARFAGARCTLHQRFSRYFTGQILATEAVYALAOQFGHDPQAMALAFVNGRP 306
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 281 GIAGAVIGPRPEQDSALKASAMTLDGALSELDEI 317
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 307 FVASNITGATTMEQLKSNLSDLSLNLNELLQIOEI 343
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
```

Search completed: June 22, 2001, 09:25:45
Job time: 110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2001, 09:24:10 ; Search time 11.8 Seconds

(Without alignments)
966,701 Million cell updates/sec

Title: us-09-463-705a-2

Sequence: 1 MTIDATHTVRLGRSALLTSR.....LDEIFPAVASGGAPEAWIQ 333

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	418	24.2	324	1 YAJQ_ECOLI	P77735 escherichia
2	334	19.3	346	1 TAS_ECOLI	O46933 escherichia
3	312	18.1	387	1 STCV_EMERI	O00727 emeticella
4	307	17.6	307	1 A115_TOBAC	P40691 nicotiana t
5	303.5	17.6	310	1 YCCB_BACSU	P46905 bacillus su
6	293.5	17.0	331	1 G569_BACSU	P80874 bacillus su
7	286	16.6	310	1 IOLB_BACSU	P46336 bacillus su
8	285	16.5	306	1 IN22_MAIZE	P49249 zea mays (m
9	285	16.5	388	1 NORA_ASPEL	O00049 aspergillus
10	283	16.4	388	1 NORA_ASPEL	O00258 aspergillus
11	273	15.8	376	1 AADL_YEAST	P48884 saccharomyc
12	268	15.5	385	1 AADL_YEAST	O01752 phanerochaete
13	263.5	15.2	329	1 AADL_YEAST	O01747 saccharomyc
14	256.5	14.8	340	1 YAKC_SCHPO	O09923 schizosacch
15	255.5	14.8	306	1 YOKF_BACSU	P54569 bacillus su
16	255.5	14.8	363	1 AAD3_YEAST	P25612 saccharomyc
17	253.5	14.7	323	1 YW98_MYCTU	O50668 mycobacteri
18	244.5	14.1	300	1 YCSN_BACSU	P49972 bacillus su
19	244.5	14.1	326	1 YDUG_ECOLI	P72556 escherichia
20	217.5	12.6	330	1 AR72_HUMAN	O44688 homo sapien
21	203.5	11.8	285	1 ALKE_RABBO	P40690 haresa bovin
22	202.5	11.7	331	1 AR72_HUMAN	O95154 homo sapien
23	198	11.5	288	1 AADA_YEAST	P47182 saccharomyc
24	196.5	11.4	327	1 AR71_RAT	P38918 rattus norv
25	184	10.6	267	1 YAFB_ECOLI	P30863 escherichia
26	169	9.8	298	1 YDHF_ECOLI	P76187 escherichia
27	168.5	9.6	286	1 YDHC_ECOLI	P25905 escherichia
28	162	9.4	284	1 P100_LEIMA	P22045 leishmania
29	161.5	9.3	318	1 XYL1_PICST	P31867 picchia stip
30	158	9.1	275	1 YDHE_ECOLI	O46857 escherichia
31	157	8.9	277	1 ZDKG_CORSP	P15339 corynebacte
32	153.5	8.9	324	1 ALDX_RAT	P15635 rattus norv
33	152.5	8.6	312	1 YPR1_YEAST	O12458 saccharomyc

34	152.5	8.8	324	1 ALDX_PIG	P50578 sus scrofa
35	150.5	8.7	295	1 MORL_PSEPU	O02198 pseudomonas
36	149.5	8.7	324	1 ALDX_HUMAN	P14550 homo sapien
37	149	8.6	310	1 S6PD_MALDO	P28475 malus domes
38	148.5	8.6	320	1 ALDR_HORVU	P23901 hordeum vul
39	147	8.5	321	1 YDGT_SCHPO	O10494 schizosacch
40	144.5	8.4	312	1 GCY_YEAST	P14065 saccharomyc
41	132.5	7.7	282	1 Y166_YEAST	P47137 saccharomyc
42	129.5	7.5	103	1 YCR4_ERWHE	O01333 erwinia her
43	127	7.3	318	1 XYL1_PACPA	P78736 pachysolen
44	125.5	7.3	329	1 XYL1_KLUFA	P49378 kluyveromyc
45	125	7.2	212	1 AAD6_YEAST	P43547 saccharomyc

ALIGNMENTS

```

RESULT 1
YAJQ_ECOLI STANDARD: PRT: 324 AA.
ID YAJQ_ECOLI
AC P77735:
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN PGPA-1SPA INTERGENIC REGION.
GN YAJQ.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
[2]
RN [2]
RP SEQUENCE FROM N.A.
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federspiel N., Hyman R., Kalman S., Komp C., Kurd O.,
RA Lew H., Lin D., Namath A., Oether P., Schramm S., Davis R.W.;
RL Submitted (Jun-1997) to the EMBL/GenBank/DBJ databases.
CC - SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to: license@sib-sib.ch).
CC
CC EMBL: AE000148; AAC73522.1; ALT_INIT.
CC EMBL: U82664; AAB40175.1; ALT_INIT.
CC EMBL: EG13611; YAJQ.
CC Hypothetical protein: Oxidoreductase.
CC ACT SITE 127 HYDROGEN-BOND DONOR (PROBABLE).
CC SEQUENCE 324 AA: 36420 MW: 0869E09F2BED9B1 CRC64:

```

Query Match 24.2%; Score 418; DB 1; Length 324;
Best local similarity 34.6%; Pred. No. 8e-26;
Matches 112; Conservative 52; Mismatches 124; Indels 36; Gaps 7;

```

OY 11 LGRSALLTSRLVGTVNF-----SGRVEDDRLRLMDHARDGINCLOTDADYGMWL 62
|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 6 LKKTDLRVSRLCLGCMFGCEPDGRGNHAWTLPESSRPITKRALEGINFEDTANSYS--- 62
OY 63 YGHTHEELVGRVLAAGCGGRREDTVALTKVGGMSERVND--SGLSARHTIASCEGSLRL 120

```

```

Db      63  -DGSSEIYGRAL-RDPARREDVYATKV-----FHRVGDLPESLSRAQLIRSIDSLRRL 116
OY      121 GVDHIDVYOMHIDR-APDEWQAMDSLVASQSVYSSNAGNHIAAOENARRHS 180
Db      117 GMDYVYDIIHREDYTPLEETLEALNDVYKAKARVIGASSMNASQFOALELOROHGW 176
OY      181 LGVSHQCYLNLAVRRAEVLPLPAOAGYGFVWSPHGLSLG-----ALEKLA 231
Db      177 AGVSNQMDHYNLIYREEREMLPICVQEGVAVIPMSPLARGRLTRRWGETTARLVSDVGC 236
OY      232 AGTAVKSAOQAVLLPSLRPAIEAVEKFCRNIGEDPAEYGLAWLSPGICAGAVIGPRT 291
Db      237 KNLKRESDEMDAQI-----AERLTGVSEELGATRAQVAILMLSKPGIAPITGSTR 288
OY      292 PEQDLSAKASAMTLDQALSELD 315
Db      289 EROJDELINAVDITTLKPEQIAELE 312

```

RESULT 2

```

TAS_ECOLI
ID TAS_ECOLI STANDARD: PRT: 346 AA.
AC 046833:

```

```

DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TAS PROTEIN.
GN TAS.

```

```

OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.

```

```

RX NCBI_TaxID=562;

```

```

RC STRAIN=K12 / MG1655;

```

```

RX MEDLINE=97426617; PubMed=9278503;

```

```

RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;

```

```

RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).

```

```

RN [12]

```

```

RC SEQUENCE FROM N.A.

```

```

RX MEDLINE=98240727; PubMed=9560382;

```

```

RA Timms A.R., Bridges B.A.;

```

```

RT "Reversion of the tyrosine ochre strain Escherichia coli WU3610 under
RT starvation conditions depends on a new gene tas."

```

```

CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

```

CC EMBL: 029561; AAB0481.1;

```

```

DR EMBL: AE000367; AAC75873.1;

```

```

DR HSSP: P14550; ZALR.

```

```

DR Ecocyc: ECG13093; tas.

```

```

DR InterPro: IPR001395;

```

```

DR Pfam: PF00248; aldo_ket_red; 1.

```

```

FT Oxioreductase.
KW ACT SITE 132 132 HYDROGEN-BOND DONOR (BY SIMILARITY).
SQ SEQUENCE 346 AA; 38499 MW; 111692D06CA07CD7 CRC64;

```

```

Query Match 19.3%; Score 334; DB 1; Length 346;
Best Local Similarity 30.4%; Pred. No. 3.4e-19;
Matches 106; Conservative 62; Mismatches 135; Indels 46; Gaps 10;

```

```

OY      10 RIGRSALLSRMLGTGVNSGRVYEDDALARLMDHARDGICLDTADMY----GWRLYKQH 66
Db      5 RPHSSLEVSTLGLTGMTGEONSEADAHQOLDYAAOAGINLIDVEMVPPVPPETQGL 64
OY      67 TEELVGRWLAOGGRRDPTVLATKVGEMSERVNDG-----LSRHHIASCEGLRL 120
Db      65 TETVYGNMLAKHGS-REKLIISKVSG--PSRNDGCIIPDQALDKNNIREALHDSLKR 121
OY      121 GVDHIDVYOMHIDR-----SAPWDEWQAMDSLV--ASQKSVYSSNF 163
Db      122 QYDYLDTLVGVHMPQRPNTCEGKLGYSMTDSAPAVSLDTLDLALAEYQRAKIRYIGVNE 181
OY      164 AGNHIAAOENARRHSLSGVSHQCYLNLAVRRAEVLPLPAOAGYGFVWSPHGLSL 223
Db      182 TAFGVMRVYHLADKHDLPRTVITQNPYSLNRSFEGVLAESQYEGVELLAISGLGFTL 241
OY      224 SGALERKLAAGTAVKSAOQAVLLPSLR-----PAIEAVEKFCRNIGEDPAEYGLAW 276
Db      242 TG---KYLNG--AKPAGARNTLFSTRYSGEOTQAKAVAVYDIARRHGLDPAQMALAV 296
OY      277 LSRPGIAGAVIGPRTPEQDLSAKASAMTLDQALSELDI-----FPA 320
Db      297 RROPFVASTLLGATMTDQKTNIESLHLESEDLVAIEAVHQVYTPA 345

```

RESULT 3

```

ID STCV_EMENT STANDARD: PRT: 387 AA.
AC 000727:

```

```

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 38, Last annotation update)

```

```

DE PUTATIVE STERIGMATOCYSTIN BIOSYNTHESIS DEHYDROGENASE STCV
DE (EC 1.1.1.-).

```

```

GN STCV.

```

```

OS Emeritella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiiales; Trichocommataceae; Emeritella.

```

```

OX NCBI_TaxID=5072;

```

```

RN [11]

```

```

RC SEQUENCE FROM N.A.

```

```

RX MEDLINE=96202293; PubMed=8643646;

```

```

RA Brown D.W., Yu J.-H., Kelkar H.S., Fernandes M., Nesbitt T.C.,
RA Keller N.P., Adams T.H., Leonard T.J.;

```

```

RT "Twenty-five coregulated transcripts define a sterigmatocystin gene
RT cluster in Aspergillus nidulans."

```

```

RL Proc. Natl. Acad. Sci. U.S.A. 93:1418-1422(1996).

```

```

CC -1- FUNCTION: MAY BE INVOLVED IN THE DEHYDRATION OF 5-
CC HYDROXYAVERANIN TO FORM AVERUPIN.

```

```

CC -1- PATHWAY: STERIGMATOCYSTIN BIOSYNTHESIS.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

```

```

CC EMBL: U34740; AAC49206.1;

```

```

DR EMBL: U34740; AAC49206.1;

```

```

DR ActOxidoreductase.
KW ACT SITE 148 148 HYDROGEN-BOND DONOR (POTENTIAL).
SQ SEQUENCE 387 AA; 43548 MW; 1621588273B85E8C CRC64;

```

```

Query Match 18.1%; Score 312; DB 1; Length 387;
Best Local Similarity 30.2%; Pred. No. 2.1e-17;

```

Matches 98: Conservative 57: Mismatches 135: Indels 34: Gaps 10:

QY 14 SALLTSRLMLGTGNFSGR-----VEDDRLMDHARDGICLDTADMYGMLYKCH 66
DB 23 AGIRSPCLGTMHFGCGMMRAMGVDVETKETAFFALLDREYVAGNFIDTAN-----FYQGE 77
QY 67 -TEEVYGRMLAOGGGRREDTVLTK-----VGGEMSERVNDGSLARHIIASCGESIR 119
DB 78 GSEKMLGEMVA-SRGNRDVLATKTYTMSYRITGPEKIKSNFGSHSLSLVASTLAK 136
QY 120 LGVDHIDVYOMHHIDRSAPMDEVMQAMDSLVAAGKVSYYGSSNFAGWHIAAOENARRH 179
DB 127 LATDYIDLLYHMMDFSPSVVEEVMQSLHVLVAGKVLNIGISDAPAVYAKNE-YARFH 195
QY 180 SIGMYSHOCILY- -LAVRAHEVLVPAQAYGICVFAMSPFHGGLSLGALEKLAAGTA 235
DB 196 GL- -TRCVYOGRWACSRPFEREILPMCOSEGLALAPWALGIRGOYKSAEEFOEGR 252
QY 236 VKSAGRAQVLLPSLRPAIEAEKFCRNLGDEPAEVLAWLVSRGVAGVAGPTPEQL 295
DB 253 NMGPQOEKRLMGA--KLTVEGER- -KGVAALALALVLLHKSPPYFVPIGCRTEQL 306
QY 296 DSALKASAMTDEQALSELDEIFP 319
DB 307 EANITSLGVELSDEIYEIEDTIP 330

RESULT 4
A115.TOBAC STANDARD: PRT: 307 AA.
AC P40691:

DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE AUXIN-INDUCED PROTEIN PCNT115.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota: Viridiplantae: Embryophyta: Tracheophyta: Spermatophyta;
OC Magnoliophyta: eudicotyledons: core eudicots: Asteridae: euasterids I;
OC Solanales: Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV WHITE BURLLEY;
RA van der Zaai E.J.;
RL Submitted (AUG-1990) to the EMBL/GenBank/DBJ databases.
CC CC
CC -1- INDUCTION: BY AUXIN.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC CC
CC DR EMBL; X56267; CA39708.1; -
CC DR PIR; S16390; S16390.
CC DR InterPro: IPR001395; -
CC DR Pfam; PF00248; aldo_ket_red; 1.
CC KW Oxidoreductase.
CC FT ACT_SITE 136
CC SEQUENCE 307 AA; 33857 MW; C4417DA852613360 CRC64;

Query Match 17.8%; Score 307; DB 1: Length 307;
Best Local Similarity 30.1%; Pred. No. 3.9e-17;
Matches 92: Conservative 50: Mismatches 136: Indels 28: Gaps 8;

QY 9 VILGRSALLTSRLMLGTGNFSG- -RVEDDRLMDHARDGICLDTADMYGMLYK 64
DB 11 IRLSGGLEVAAGLGGMGMSAFYGPAPKPEPMIQLIHAIINSGITLDTSVYQ- -65

QY 65 GHTEE-LVGRMLAOGGGRREDTVLTKVGGEMSERVNDGSLARH- -IIASCEGSIR 118
DB 66 PHTNELLLGR- -ALGGTRERVLATKFGIYLGDCKAKGKAVGDPAYVAACEASLK 123
QY 119 RGVGDHIDVYOMHHIDRSAPMDEVMQAMDSLVAAGKVSYYGSSNFAGWHIAAOENARRH 178
DB 124 RLDIDICIDLYHRDTRPFIETVYLGELKKLYBECKLAKYIGSE- -ASASTIRRA 176
QY 179 HSLGAVSH-QCLYNLAVRAHEVLVPAQAYGICVFAMSPFHGGLSLGALEKLAAGTA 237
DB 177 HAVHPIITAVQLTWSLMSRDEVEEIIPTCRELGIGIYAYSPLRGFLSSGPKLEDMNSND 236
QY 238 SAQGRAQVLLPSLRPAIEAEKFCR- -NLGSDPAEVLAWLVSRGVAGVAGPTPEQL 294
DB 237 YRKYLPFRQAEMLNKKLYERICEAVARKGCTPSOLALAWHOGQNDVCPITGTTKEN 296
QY 295 LDSALK 300
DB 297 LNMNMK 302

RESULT 5

YCKK_BACSU STANDARD: PRT: 310 AA.
ID YCKK_BACSU

AC P46905:
DT 01-NOV-1995 (Rel. 32, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN NATB-RAPD INTERGENIC REGION (ORF).
GN YCKK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kuno M., Tamakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation".
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-268 FROM N.A.
RC STRAIN=BD99;
RA Cheng J., Guffanti A.A., Krulwich T.A.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC CC
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC CC
CC DR EMBL; AB000617; BAA22238.1; -
CC DR EMBL; Z99105; CAB12071.1; -
CC DR EMBL; U30873; AAB53024.1; -
CC DR HSSP; P80276; IAH3.
CC DR Subtilist; BG11340; YCKK.
CC DR InterPro: IPR001395; -
CC DR Pfam; PF00248; aldo_ket_red; 1.
CC KW Hypothetical protein; Oxidoreductase.
CC FT ACT_SITE 127
CC SEQUENCE 310 AA; 34800 MW; C06BF4195D25C91C CRC64;

Query Match 17.6%; Score 303.5; DB 1: Length 310;
Best Local Similarity 28.9%; Pred. No. 7.4e-17;
Matches 92: Conservative 54: Mismatches 147: Indels 25: Gaps 8;

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AB005554; BAA21607.1; -
 DR EMBL; 299124; CAB16014.1; -
 DR HSSP; P23457; LMI.
 DR Subtilist; BG11363; 101S.
 DR Pfam; PF00246; aldo_ket_red; 1.
 DR Oxidoreductase.
 FT ACT_SITE 125 125 HYDROGEN-BOND DONOR (PROBABLE).
 SO SEQUENCE 310 AA; 35168 MW; A870F226F8684867 CRC64;

Query Match 16.6%; Score 286; DB 1; Length 310;
 Best Local Similarity 28.6%; Pred. No. 1.8e-15;
 Matches 95; Conservative 58; Mismatches 129; Indels 50; Gaps 11;

QY 10 RLGSALLTSLRMIGTVNFGS-----RVEDDALRLMDHARDRGINCIDTADMYGRLYK 64
 Db 5 KLGSQGLVSAQGGCGMGSFAFGPPRESEMRKLHADVAGVTFDTSVYV-----65
 QY 65 GHTSELVGRMLAOGGREDTVALTKVG--GEMSERVNDLSARHIIASCEGSLRLGVD 123
 Db 61 GRSELLEGEVIRE--FNEDEVIAAKAHRKOGNDEVEDN--SPDEFLKSVDESILKRLNT 116
 QY 123 DHIVYQMHIDRSAPWDEWQAMDSLVAAGKVSYSVSSNFGWHIAAQAENARRHSLG 182
 Db 117 DYIDLFYHFDEHPKQEAVALNEMKAKKIRSIGVSNFSLDKRANND-----G 169
 QY 183 MVS-HQCLYNLAVRHALEVLPAQAQYGLGVFAMSPFHGSLGALAKTAVKASAG 241
 Db 170 LVDLVQGEYNLNREAEKTFEPYKRNHISFIPYFPLVSGLAG--KYTEDTFEPES-- 224
 QY 242 RAQVLLPRLRAIEAY--EKFCRNIGE-----DPAVEGLAWVLSRQIAGAV 286
 Db 225 -----DLRDEQHEFKERKERNIRKYNKLAFLAEKHNVDIPHLAWYLARPEIDILI 277
 QY 287 IGPRTEQDLSALKASAMTIDEQALSELDEIF 318
 Db 278 PGAKRADQLINIKTADVTLSQEDISFIDKLF 309

RESULT 8
 IN22_MAIZE STANDARD; PRT; 306 AA.
 AC P49249;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE IN2-2 PROTEIN.
 GN IN2-2 OR SAF2.
 OS Zea mays (Maize).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;
 OC Andropogoneae; Zea.
 OX NCBI_TaxID=4577;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV_MISSOURI 17;
 RX MEDLINE=92003682; PubMed=1912492;
 RA Hershey H.P., Stoner T.D.;
 RT "Isolation and characterization of cDNA clones for RNA species
 RT induced by substituted benzene sulfonamides in corn.";
 RL Plant Mol. Biol. 17:679-690(1991).
 CC -1- TISSUE SPECIFICITY: LEAF AND ROOT.
 CC -1- DEVELOPMENTAL STAGE: IT APPEARS IN ROOTS WITHIN 30 MIN OF
 CC INDUCTION, MAXIMUM LEVELS ARE REACHED BY 6 HRS, AND REMAINS
 CC CONSTANT FOR 2 DAYS. IN LEAVES IT IS SEEN 9 HRS AFTER INDUCTION,
 CC AND REACHES MAXIMUM LEVELS AFTER 24 HRS.
 CC -1- INDUCTION: BY N-(AMINOCARBONYL)-2-CHLOROBENZENESULFONAMIDE (2-

CC (CBSU).
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
 CC MaizeDB; 121988; -
 DR Oxidoreductase.
 FT ACT_SITE 131 131 HYDROGEN-BOND DONOR (PROBABLE).
 SO SEQUENCE 306 AA; 33828 MW; 2E94674CFDE3C8D CRC64;

Query Match 16.5%; Score 285; DB 1; Length 306;
 Best Local Similarity 32.9%; Pred. No. 2.1e-15;
 Matches 73; Conservative 40; Mismatches 89; Indels 20; Gaps 6;

QY 9 VRLGSALLTSLRMIGTVNFGS-----RVEDDALRLMDHARDRGINCIDTADMYGRLYK 64
 Db 11 KLGSQGLVSAQGGCGMGSFAFGPPRESEMRKLHADVAGVTFDTSVYV-----65
 QY 65 GHTSELVGRMLAOGGREDTVALTKVGEMSERVNDLSARHIIASCEGSLRLGVD 123
 Db 66 PHTNVLVIGK--ALOGGVREKVELATKFGVSFADGKREIHGDPAYVTRACGSKRLGVD 123
 QY 124 HIDVYQMHIDRSAPWDEWQAMDSLVAAGKVSYSVSSNFGWHIAAQAENARRHSLG 183
 Db 124 CIDLYQHRIDKRVPIETVIGELKTLVEGKIKYIGLSE-----ASASTIRRAHAYHP 176
 QY 184 VSH-QCLYNLAVRHALEVLPAQAQYGLGVFAMSPFHGSLG 224
 Db 177 ITAVOLEWMSLMSRDAEEDIIPTCRELIGIGIAYVSLGGRFES 218

RESULT 9
 NORA.ASPFL STANDARD; PRT; 388 AA.
 AC Q00049;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE NORSOLORINIC ACID REDUCTASE (EC 1.1.1.1-).
 GN NORA OR ADH-2.
 OS Aspergillus flavus.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5059;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL 3357;
 RX MEDLINE=96156784; PubMed=8593042;
 RA Cary J.W., Wright M., Bhatnagar D., Lee R., Chu F.;
 RT "Molecular characterization of an Aspergillus parasiticus
 RT dehydrogenase gene, nora, located on the aflatoxin biosynthesis gene
 RT cluster".
 RL Appl. Environ. Microbiol. 62:360-366(1996).
 CC -1- PATHWAY: AFLATOXIN BIOSYNTHESIS.
 CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; U32377; AAC49167.1; -
 DR Oxidoreductase.
 FT ACT_SITE 148 148 HYDROGEN-BOND DONOR (POTENTIAL).
 SO SEQUENCE 388 AA; 43751 MW; 758B8187187BF580 CRC64;

Query Match 16.5%; Score 285; DB 1; Length 388;
 Best Local Similarity 28.6%; Pred. No. 2.8e-15;
 Matches 97; Conservative 58; Mismatches 132; Indels 52; Gaps 12;

QY 8 HVRLGRSA-LITSRLMIGTVNFS-----GRVEDDALRLMDHARDRGINCIDTADMYG 59

```

Db      16 HLLSSASAVRSPCLGLTMSFGNGKVGWCECDQATSFNMIDTFEYSGGNFIDVANYF- 74
Oy      60 WRLYKCHTEELVGRMLAOGGGRREDVLA TK-----VGGEMSERVNDGSLARHIIASG 113
Db      75 ---OGGDTERWVGEMWAAQ-RQNDEIVLSTKYTMGYMFPKIRKSNFQGNHAKSLRLSV 130
Oy      114 EGSRLRLGVADHDIVYOMHIDRSAPMDEWQAMDSLVAASKVSYGSSNFAWHIAAOE 173
Db      131 KASLQKLTQTDYIDLVLVHMDDFTTSVEEYMRSLNHLVANGKVLILGVSDTPALVYKC-- 188
Oy      174 NA-ARRHSLGWSV-HQCLYNLAVRHALEVLPAQAAYGLGVFAMSPDLHGGLSGALEKLA 231
Db      189 NAFARANGILTPEVSVOGHWSAFRDEFDILPWCSESEGMGLAPWGLVGRGQFRSAEE--- 245
Oy      232 AGTAVSAGRAQVLLPSLRPAIEAEKFCRNIG-----DPAEVLAWLNSRP 280
Db      246 -----FSRGR-----KKGPODEKH-----RLGEKLDQMAOQKNTKATSIQAIVMHKA 290
Oy      281 GIAGAVIGRTEPQDLSALKASAMTLEQALSELDEIFP 319
Db      291 PYEPIYIGRKVEHLKENIEALGLVLSSEEIREIDAEP 329

```

```

RESULT 10
NORA_ASPPA STANDARD: PRT: 388 AA.
AC 000258:
DT 15-JUL-1999 (rel. 38, Created)
DT 15-JUL-1999 (rel. 38, Last sequence update)
DE 15-JUL-1999 (rel. 38, Last annotation update)
DE NORSOLOINIC ACID REDUCTASE (EC 1.1.1.-).
GN NORA.
OS Aspergillus parasiticus.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5067;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 163 / NRRL 5862 / SU-7;
RX MEDLINE=96156784; PubMed=8593042;
RA Cary J.W., Wright M., Bhattacharjee D., Lee R., Chu F.;
RT "Molecular characterization of an Aspergillus parasiticus
RT dehydrogenase gene, nora, located on the aflatoxin biosynthesis gene
RT cluster."
RT Appl. Environ. Microbiol. 62:360-366(1996).
RL Appl. Microbiol. 62:360-366(1996).
CC -1- PATHWAY: AFLATOXIN BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U24698; AAC49166.1; -.
CC OXidoreductase.
CC FT ACT_SITE 148 HYDROGEN-BOND DONOR (POTENTIAL).
CC SEQUENCE 388 AA; 43721 MW; 8829491E6E6F6BB CRC64;

```

Query Match 16.4%; Score 283; DB 1; Length 388;
 Best Local Similarity 28.6%; Pred. No. 4e-15;
 Matches 97; Conservative 58; Mismatches 132; Indels 52; Gaps 12;

```

Oy      8 HAVLGSALTLTSLRLGTVNFS-----GRVEDDALRLMDHARDGNCIDLTADMG 59
Db      16 HRLSSASAVRSPCLGLTMSFGNGKVGWCECDQATSFNMIDTFEYSGGNFIDVANYF- 74
Oy      60 WRLYKCHTEELVGRMLAOGGGRREDVLA TK-----VGGEMSERVNDGSLARHIIASG 113

```

```

Db      75 ---OGGDTERWVGEMWAAQ-RQNDEIVLSTKYTMGYMFPKIRKSNFQGNHAKSLRLSV 130
Oy      114 EGSRLRLGVADHDIVYOMHIDRSAPMDEWQAMDSLVAASKVSYGSSNFAWHIAAOE 173
Db      131 KASLQKLTQTDYIDLVLVHMDDFTTSVEEYMRSLNHLVANGKVLILGVSDTPALVYKC-- 188
Oy      174 NA-ARRHSLGWSV-HQCLYNLAVRHALEVLPAQAAYGLGVFAMSPDLHGGLSGALEKLA 231
Db      189 NAFARANGILTPEVSVOGHWSAFRDEFDILPWCSESEGMGLAPWGLVGRGQFRSAEE--- 245
Oy      232 AGTAVSAGRAQVLLPSLRPAIEAEKFCRNIG-----DPAEVLAWLNSRP 280
Db      246 -----FSRGR-----KKGPODEKH-----RLGEKLDQMAOQKNTKATSIQAIVMHKA 290
Oy      281 GIAGAVIGRTEPQDLSALKASAMTLEQALSELDEIFP 319
Db      291 PYEPIYIGRKVEHLKENIEALGLVLSSEEIREIDAEP 329

```

```

RESULT 11
AADE YEAST STANDARD: PRT: 376 AA.
AC P42884:
DT 01-NOV-1995 (rel. 32, Created)
DT 01-NOV-1995 (rel. 32, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE PUTATIVE ARYL-ALCOHOL DEHYDROGENASE AAD14 (EC 1.1.1.-).
GN AAD14 OR YNL331C OR N0300.
OS Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RA Van Dyck L., Pascual-Anhui A., Goffeau A.;
RT Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RC Omermaier B., Piravandi E., Rinke M.;
RT Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
RL [2]
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY. STRONG,
CC TO YEAST AAD6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X83226; CAAS8227.1; -.
CC DR EMBL: Z71607; CAAS6264.1; -.
CC SCD: S0005275; AAD14.
CC OXidoreductase.
CC FT ACT_SITE 151 HYDROGEN-BOND DONOR (POTENTIAL).
CC SEQUENCE 376 AA; 41991 MW; 2265406386938313 CRC64;

```

Query Match 15.8%; Score 273; DB 1; Length 376;
 Best Local Similarity 27.0%; Pred. No. 2.3e-14;
 Matches 86; Conservative 51; Mismatches 142; Indels 40; Gaps 8;

```

Oy      22 WLGTVNFSGRVEDDALRLMDHARDGNCIDLTADMGWRLYKCHTEELVGRMLAOGGR 81
Db      43 WSG---FMGSMNKEQAFELIDAFYFAGGNCIDTANYSY-----QNESEIWIWEMWA-SRKL 94
Oy      82 REDTVLATATVGGMSFRVNDGSLA-----RHIIASCEGSLRLGVADHDIVYOMHHHD 134
Db      95 RQDIVLATATFTGDKRYVEVGSGKSNVCGNKRSLHVSVDSTKLTQTDYIDLVLVHMDD 154
Oy      135 RSAPMDEWQAMDSLVAASKVSYGSSNFAWHIAAOENAAARRHSLGWSHQCLYNLA 194

```

```

Db 155 YMSSTEEWMDSHLIVQCKVLLTVSDTPAWVVSAAANYATSHGKTPPSYQGMWNLN 214
Oy 195 RHAEELEVPAAQAYGLGYFAWSPHLGGLSGALEKLAAGTAVKASQGRAQVLLPRLPAI 254
Db 215 RDEFDITIPMARHFGMALAPMDVMGGRFQ---SKKAMEERKKNES-----LRFV 263
Oy 255 -----EAYEFCRNMG-EDPAEVLAVLSRPGIAGVIGPRPEQLDLSLK 300
Db 264 GGEPEETELEVKSALTETIAEEHGTESVTALAIAYRSKAKNVFLLIGRKIEHLKONIE 323
Oy 301 ASAMTLDQALSELDEIFP 319
Db 324 ALSTKLPPEQILEYESTVP 342

RESULT 12
AAD_PHACH STANDARD: PRT: 385 AA.
AC 001752.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ARYL-ALCOHOL DEHYDROGENASE [NADP+] (EC 1.1.1.91) (AAD).
OS Phanerochaete chrysosporium.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Aphylllophorales;
OC Corticiaceae; Phanerochaete.
OX NCBI_TaxID=5306;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-BKM-F-1767;
RX MEDLINE=95050593; PubMed=7961751;
RA Reiser J., Wehlem A., Hardegger M., Frank G., Fiechter A.;
RT "Aryl-alcohol dehydrogenase from the white-rot fungus Phanerochaete
chrysosporium. Gene cloning, sequence analysis, expression, and
RT purification of the recombinant enzyme.";
RL J. Biol. Chem. 269:28152-28159(1994).
CC -I- CATALYTIC ACTIVITY: AN AROMATIC ALCOHOL + NADP(+) = AN AROMATIC
CC ALDEHYDE + NADPH.
CC -I- PPM: THE N-TERMINUS IS BLOCKED.
CC -I- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; L08964; AAA61931.1; -
RW Aromatic hydrocarbons catabolism; Oxidoreductase; NADP.
FT ACT_SITE 153 153 HYDROGEN-BOND DONOR (POTENTIAL).
SQ SEQUENCE 385 AA; 43563 MW; 7D5DEB0690B9A20 CRC64;

Query Match 15.5%; Score 268; DB 1; Length 385;
Best Local Similarity 24.0%; Pred. NO. 5.9e-14;
Matches 82; Conservative 73; Mismatches 147; Indels 40; Gaps 8;

Oy 10 RLGR-----SALTSLRMIGTVNFS-----GRVEDDALRLMDHARDRGINCIDT 54
Db 13 KLGRHROLAPGCCGLHVSITQLGAMSTGDKMHPYGMGTMDKESFLDLDAFYVAGNFTDT 72
Oy 55 ADMYGRLYKGTTELTVRWLAOGGGRREDTVLATKVG-----GEMSERVNDGSL 105
Db 73 ANVY-----QDESEEFIEEMW-EARGNRDQWVAATKYSLYVKRGASFEIEPKOTYVENS 127
Oy 106 ARHITASEGSLRLGLVDHIDVYQMHHTDRSAPMDEVWQAMDSLVAASKVSYGSSNPA 165
Db 128 LKSMHISVHDSLRKLRYSIDIFVYHFWDYCTTIEVVMKGNHLNLAOGKVLVLYGSDTPA 187
Oy 166 WHIAAOGENAARRHSLGVNSHQCCLYNLAIRHALELEVLPAAQAYGLGYFAWSPHLGGLSG 225

```

```

Db 188 WYVSKANNYARNAGKTPPEVTEGEMNITWRDMEKDIIPICCHEGMAIPWNLCAKGIKRT 247
Oy 226 ALEKLAAGTAVKASQGRAQVLL-----PSLRPIAEAYEKFCRNMG-EDPAEVLAVL 277
Db 248 DAEE---ERRLKSQGGGRTLLQFDGMLNRETERRKSKALEKAEELGAKSITSVAIALYM 304
Oy 278 SRPGIAGVIGPRPEQLDLSALKASAMTLDQALSELDEIFP 319
Db 305 QKFPYVFPVIGGRKVEHLYANLEALDISPQOMQFLNDTVP 346

RESULT 13
AAD4_YEAST STANDARD: PRT: 329 AA.
AC 007747.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE PUTATIVE ARYL-ALCOHOL DEHYDROGENASE AAD4 (EC 1.1.1.-).
GN AAD4 OR YDL243C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Alt-Moerbe J., Schneider C., Moro M.;
RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; 274291; CAA98823.1; -
RW Oxidoreductase.
FT ACT_SITE 105 105 HYDROGEN-BOND DONOR (POTENTIAL).
SQ SEQUENCE 329 AA; 36977 MW; 62BDBD0C889610E0 CRC64;

Query Match 15.2%; Score 263.5; DB 1; Length 329;
Best Local Similarity 25.6%; Pred. NO. 1.1e-13;
Matches 77; Conservative 60; Mismatches 147; Indels 17; Gaps 6;

Oy 30 GRVEDDALRLMDHARDRGINCIDTADMYGRLYKGTTELTVRWLAOGGGRREDTVLAT 89
Db 2 GSMNEQAFELDLDAFYEAGNCIDITRANSY---QNESEIWIWEMW-KSRKLRDQIVAT 56
Oy 90 KYVGEMSERVNDGSLSAH-----IIASCGSLRLGLVDHIDVYQMHHTDRSAPMDEV 142
Db 57 KFTGQKRYEYVGGGSKSNVYCGNKHKSLHVSYSRLKLTQTDIDILYHMMWYMSISIEV 116
Oy 143 WQAMDSLVAASKVSYGSSNPAFWHIIAAOENAAARRHSLGVNSHQCCLYNLAIRHALEVL 202
Db 117 MDSLHILVQGGKVLVLYGSDTPAWVVSAAANYATSHGKTPPSYQGMWNLNLRDEKROI 176
Oy 203 PAAQAYGLGYFAWSPHLGGLSGAL--EKLAAGTAVKASQGRAQVLLPSLRPAIAYEK 259
Db 177 PWARHFGMALAPMDVMGGRGFQSKRAMBERKKNKGSLRTVSGTSKQTPKEVXIS-EALAK 235
Oy 260 FCNRNG-EDPAEVLAVLSRPGIAGVIGPRPEQLDLSALKASAMTLDQALSELDEIFP 318
Db 236 VAEEHGTESVTALAIAYRSKAKNVFVLGGRKIEHLKONIEALSIKLPPEQILEYESTI 295
Oy 319 P 319
Db 296 P 296

```

```

RESULT 14
YAKC_SCHPO STANDARD: PRT: 340 AA.
AC 009923;
DT 01-FEB-1996 (rel. 33, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DE 01-FEB-1996 (rel. 33, Last annotation update)
DE HYPOTHEMETICAL 37.7 KDA PROTEIN C1F7.12 IN CHROMOSOME 1.
GN SPAC1F7.12 OR SPAC21E11.01.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Murphy L., Niblett D., Harris D., Barrell B.G., Rajadream M.A.,
RA Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 267998; CAA91959.1; -
CC DR EMBL: 267999; CAA91961.1; -
CC DR HSSP: P14550; 2ALR.
CC DR InterPro: IPR001395; -
CC DR Pfam: PF00248; aldo_ket_red.1.
CC KW Hypothetical protein; Oxidoreductase.
CC FT ACT_SITE 126 HYDROGEN-BOND DONOR (PROBABLE).
CC FT SITE 126
CC SEQUENCE 340 AA; 37710 MW; A4A9686A70968F06 CRC64;

Query Match 14.8%; Score 256.5; DB 1; Length 340;
Best Local Similarity 26.4%; Pred. No. 4.1e-13;
Matches 78; Conservative 55; Mismatches 123; Indels .39; Gaps 9;

QY 40 LMDHARDGNCIDTADMYGWRLYKGTTELVLGRMLAOGGREDIVLATKVGEMSERV 99
DB 38 VLTHADLIGCTFMDSSDMTG-----GANEBCIRWFKO-TGRKKEFLATKFGYEKNPET 92
QY 100 NDSGLSAR--HITASCEGSLRLGVHDVYOMHIDRSAPMDEWQAMDLSVASKVSY 157
DB 93 GELSLNEDDYIEKALDLSIKRLIGDIDLYHYRFSGEPIEKINGALCKVEACKIRY 152
QY 158 VGSNNAGNHIAAOENARRHSLGWVS-HQCLYNLAVRAELEVPAAQAYLGFAMS 216
DB 153 IGLSECS---AMTIRRAAVYPVSAVOYEYSPFSLIERPEIGWAKACRENNITVCYA 208
QY 217 PLHGGLSGALEKIAAGTAVKSAQRAQVLPRLPAIEAYEK--FCRNL----- 264
DB 209 PLGRGLTG-----AYKSPDPE---GDFRKARPRYQENENYKNIELVTKIEKTA 256
QY 265 ---GEDPAEVLAWLSPRGIAAGVIGPRTPEOLDASALKASAMTIDEALSELDE 316
DB 257 TANNITPGQLSLAWLALAGDGLIPIGTKRVKYLEENFGALKVLSDAIVYKEIRE 311

```

```

DT 15-JUL-1998 (rel. 36, Last annotation update)
DE HYPOTHEMETICAL OXIDOREDUCTASE IN ANSR-BMRU INTERGENIC REGION.
GN YOKF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA Sato T., Takeuchi M.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ALDO/KETO REDUCTASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D84432; BA12638.1; -
CC DR EMBL: 299116; CABI4294.1; -
CC DR Subtilist; BG11761; YOKF.
CC DR InterPro: IPR001395; -
CC DR Pfam: PF00248; aldo_ket_red.1.
CC KW Hypothetical protein; Oxidoreductase.
CC FT ACT_SITE 119 HYDROGEN-BOND DONOR (PROBABLE).
CC FT SITE 119
CC SEQUENCE 306 AA; 34717 MW; CD79B13B956B00F CRC64;

Query Match 14.8%; Score 255.5; DB 1; Length 306;
Best Local Similarity 29.7%; Pred. No. 4.3e-13;
Matches 94; Conservative 50; Mismatches 133; Indels .39; Gaps 12;

QY 10 RIGRSALLTSRLMAGTVNFSGRVEDDALARLMDHARDGNCIDTADMYGWRLYKGTTEE 69
DB 5 KIGTSOLDISEVGLGCMISIG--TEKKKALSIIDEALIELGINVYDADLVYD---RCRNEE 58
QY 70 LVGRMLAOGGREDIVLATKVGEMSERVNDGSL-----SARHITSCEGSLRLGV 122
DB 59 IYGDALIQ---NRRHDIILITKAGNRW----DGSEGWYWDPSKRAIKAVKKSILRLKT 110
QY 123 DHTDYOMHIDRSAPMDEWQAMDLSVASKVSYGSSNFAGMHTAAQENAAARRH--S 180
DB 111 DYTIDYQLHGCGTIEDNIDETIEAFELKQEGVIRYIG-----ISSIRPNVIKRYVK 162
QY 181 LGNVSHOCLYNLAVRAELEVPAAQAYLGFVAMSPHLHGGLSGALEK-LAAGTAVKSA 239
DB 163 SNVSIIMOFSLEFDRRPE-EMLPILIEHQISVVARGPVAKGLLT---EKPLDAQSESMKO 218
QY 240 QGRAQVLLSLRALIAYKFKCNLGEDPAEVLAVLSRPGAGAVIGPRTPEQL-DSA 298
DB 219 NGLLSYSPFELTINARKAMEEVAIDL--SMTKESLOYLLAOPAVASVITGASKIEQLRENI 276
QY 299 LKASAMTIDEALSEL 314
DB 277 QAMNARRLTEEEKAL 292

```

Search completed: June 22, 2001, 09:26:00
Job time: 110 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 22, 2001, 09:23:25 ; Search time 19.19 Seconds
(without alignments)
1321.841 Million cell updates/sec

Title: US-09-463-705A-2

Perfect score: 1728

Sequence: 1 MTTPDAATVHRGSRALLTSR.....LDEIRPAVASGGAPEANLQ 333

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Query No.	Score	Match	Length	DB	ID	Description
1	804.5	46.6	329	2	E82644	sugar-phosphate de
2	791.5	45.8	362	2	D82644	sugar-phosphate de
3	763	44.2	316	2	F69978	sugar-phosphate de
4	495	28.6	354	2	T44934	moca protein (impo
5	472.5	27.3	323	2	H83427	probable oxidoredu
6	418	24.2	324	2	E85538	probable oxidoredu
7	418	24.2	348	2	C64771	probable NAD(P)H-c
8	414.5	24.0	352	2	C82294	probable oxidoredu
9	408	23.6	316	2	F83170	oxidoreductase (as
10	399	23.1	324	2	D84315	probable oxidoredu
11	393	22.7	344	2	T41659	aryl-alcohol dehy
12	384	22.2	315	2	A75289	probable potassium
13	380.5	22.0	328	2	T52133	probable potassium
14	373	21.6	345	2	B83093	potassium channel
15	371.5	21.5	342	2	S61978	probable oxidoredu
16	363	21.0	419	2	I55463	hypothetical prote
17	360.5	20.9	330	2	T07394	K+ channel beta-su
18	359.5	20.8	408	2	I59393	probable potassium
19	356	20.6	401	2	S66503	potassium channel
20	355	20.5	327	2	E75254	potassium channel
21	340	19.7	319	2	H72394	aldo/keto reductas
22	339.5	19.6	336	2	T44988	K+ channel, beta s
23	338.5	19.6	351	2	T50285	oxidoreductase (im
24	338	19.6	336	2	C84256	probable oxidoredu
25	337.5	19.5	387	2	F82815	probable oxidoredu
26	334	19.3	453	2	C65066	voltage-gated pota
27	333	19.3	453	2	C65176	hypothetical prote
28	328	19.0	367	2	A53131	hypothetical prote
29	328	19.0	367	2	S45312	alpha-dendrotoxin -
						Rckbeta2 protein -

30	326	18.9	346	2	F85935	hypothetical prote
31	325	18.8	367	2	S66502	potassium channel
32	324	18.8	329	2	C83506	probable oxidoredu
33	317.5	18.4	340	2	B96632	hypothetical prote
34	312	18.1	331	2	H83328	probable oxidoredu
35	307	17.8	307	2	S16390	auxin-induced prot
36	306	17.7	404	2	S68409	potassium channel
37	303.5	17.6	310	2	B69755	ion channel homolo
38	302.5	17.5	346	2	G65086	hypothetical prote
39	300.5	17.4	335	2	C96632	hypothetical prote
40	300	17.4	335	2	G72546	probable K+ channe
41	293.5	17.0	331	2	D69826	aldo/keto reductas
42	292.5	16.9	331	2	F85959	probable reductas
43	291.5	16.9	310	2	T35825	probable oxidoredu
44	289.5	16.8	333	2	H72307	oxidoreductase, al
45	288	16.7	338	2	T12582	auxin-induced prot

ALIGNMENTS

RESULT 1

E82644

C:Species: Xylella fastidiosa XFI724 [imported] - Xylella fastidiosa (strain 9a5c)

C:Date: 18-Aug-2000 #sequence-revision 20-Aug-2000 #text-change 02-Sep-2000

C:Accession: E82644

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MIMD:20365717

A:Note: for a complete list of authors see reference number A59328 below

A:Accession: E82644

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-329 <SIM>

A:Cross-references: GB:AE003996; GB:AE003849; NID:9106790; PIDN:AAE84533.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre

as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitzjima, J.P.; Krieger, J.E.; Kurume, E.E.; La

Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava

M.; Tsunako, M.H.; Valada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XFI724

C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 46.6%; Score 804.5; DB 2; Length 329;

Best Local Similarity 48.8%; Pred. No. 8.2e-56;

Matches 161; Conservative 52; Mismatches 106; Indels 11; Gaps 3;

QY	8	HVRGSRNLLTSRLMGLTVNFSGRVEDDRLMDHARDRCINCLDTRDMGWRK---	64
DB	3	YTHGRGLKAVSRITGLTMNGELTDEATSKIMDTALDAINFEDPADVAGGQPTDMP	62
QY	65	---GHEELGVRWLAAGGGRREDVLTATKVGSEMSERNDGSLARHITASCESLRPLG	121
DB	63	KGFSTSEYIGNWLAQDKSRDKITVLATKYQPMETGENDYLLSYHRRACEASLRK	122
QY	122	VDHIDVQMHHRIDSRAPDEWQAMDSIVASGKSYVSSNFAGMHTAAQENARRSH	181
DB	123	TDHIDVQMHHRIDSRAPDEWQAMDSIVASGKSYVSSNFAGMHTAAQENARRSH	182

Db 112 LRRLOTVDLYQSHRDDPHFDLEETLSTYGEELIKKKVAVIGASNDARLLLEAROYSA 171
 Oy 177 RRHSIGAMVSHOCL---YNLAAR-HAELEVLPAQAQVGLGVFAMSLHGGSLSGALEK--- 229
 Db 172 R---LNLPSYSGLOPEYNYLDRADYETNLEPVELGIGVISYSLAGFSGKRNQAD 228
 Oy 230 ---LAAGTAVMS-AAGRAQVLLPSLRPAIEAEKFCRNLSGEPAPVGLAMVSRPGIAGA 285
 Db 229 TAGRARGEKVKQYLNBERGVAITIAALDEVAEOY-----MANPVOYALMALIARPTTAP 281
 Oy 286 VIGPRTPEQDLSALKASAMTLDEQALSELD 315
 Db 282 IASATSLQDLDLLIATHTLKIDEQALIPED 311

RESULT 10
 D84315
 aryl-alcohol dehydrogenase [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C:Accession: D84315
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483
 A:Accession: D84315
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-324 <STO>
 A:Cross-references: GB:AE004437; NID:910581102; PIDN:AA919888.1; GSPDB:GN00138
 A:Gene: aad
 C:Genetics:
 C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 23.1%; Score 399; DB 2; Length 324;
 Best Local Similarity 35.7%; Pred. No. 5.6e-24;
 Matches 114; Conservative 40; Mismatches 143; Indels 22; Gaps 8;

Oy 1 MTDTAATVNRIGRSALLTSRLMIGTVNF-----SGRVE--DDDAKLMDHARDRCINCLD 53
 Db 1 MGLDA---VPLGRTGLRVTELAFTGFRGRETADGDIENVGQRARELLDAYANAGGRPID 57
 Oy 54 TADMGWRLYKGTETELVGRWLAQGGRRPDTVLATKVGSESEVNDGSLARHIASC 113
 Db 58 TADMGWRLYKGTETELVGRWLAQGGRRPDTVLATKVGSESEVNDGSLARHIASC 113
 Oy 114 EGSRLRLGVDHIDYVOMHIDRSAPWDEWQAMDSLVASGVSYVSSNF--AGWHIAAA 171
 Db 112 DNIITRLCTDIDILYIHRMDATAPADALMRTLDGDFVREGVYHYGASTFHPNARKIKA 171
 Oy 172 QENARHRSLSGVSHOCLYNLAARHAEVLPAQAQVGLGVFAMSLHGGSLSGALEK 231
 Db 172 NEIARRSSEPTVAQPRFNANVNEIYGNLYEMCKREGLGAVPSPSLAGFLTGKTERDA 231
 Oy 232 ---AGT-AVKSAGQRAQVLLPSLRPAIEAEKFCRNLSGEPAPVGLAMVSRPGIAGA 287
 Db 232 EPPAGSRGATQGGFVDSVLTGPSNFDTLDAVEAVADVATPAQVSLANLRHHDVAAAIT 291
 Oy 288 GPRTPPEQDLSALKASAMTL 306
 Db 292 GARTPAQLRENMAATNIEL 310

RESULT 11
 T41659
 probable potassium channel subunit - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 18-Feb-2000
 C:Accession: T41659

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
 submitted to the EMBL Data Library, May 1998
 A:Reference number: Z22008
 A:Accession: T41659
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-344 <WOO>
 A:Cross-references: EMBL:AL023590; PIDN:CAA19066.1; GSPDB:GN00068; SPDB:SPCC965.06
 A:Experimental source: strain 972h; cosmid c965
 C:Genetics:
 A:Gene: SPDB:SPCC965.06
 A:Map position: 3
 C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 22.7%; Score 393; DB 2; Length 344;
 Best Local Similarity 31.8%; Pred. No. 1.8e-23;
 Matches 107; Conservative 55; Mismatches 144; Indels 30; Gaps 10;

Oy 5 AATVNR-----LGRSALLTSRLMIG---TVNFSGRVEDDALARLMDHARDRCINCLD 53
 Db 4 ATTHFQPNVNPFRFLGRSLGKLVSAFSLGWLTYGNEG-YDVEHTRNCLKQAWDLGINTFD 62
 Oy 54 TADMGWRLYKGTETELVGRWLAQGGRRPDTVLATKV-GGEMSERVNDGSLARHIASC 112
 Db 63 TAEIYS---NNSSETVWKAKEKLGMDRSEVITTKVFFGGTXYLPNTGSLSRHILIG 118
 Oy 113 EGSRLRLGVDHIDYVOMHIDRSAPWDEWQAMDSLVASGVSYVSSNFAGWHIAAA 172
 Db 119 LNASLRLGRLPYVDYIMHRRPSPVMEVVAATFOLLQDCAFAFWGTSSEMAFELIENAH 178
 Oy 173 ENARRHSLGVSHOCLYNLAAR-HAELEVLPAQAQVGLGVFAMSLHGGSLSGAL---- 227
 Db 179 HATKYKNLIAPADQPOYNYLTFRDFEKKDLPLQOYIGYGAIVSWPSLKSGLITGKYNDGI 238
 Oy 228 -EKLAAQAVKSAQRAQVLLPSLRPAIE---AYEKFCRNLSGEPAPVGLAMVSRPGIA 283
 Db 239 PEGSRLSTFTSLAQOLPT--PEGKTOLDQVROISKIAEQIGATPSALAWTLKNPYVS 296
 Oy 284 GAVIGPRTPEQDLSALKASAM--TLDEQALSELDEI 317
 Db 297 TTIIGASKPEQIVENVKAVEFIDKLTPELKKIDDI 332

RESULT 12
 A75289
 probable potassium channel, beta subunit - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: A75289
 R:White, O.; Eissen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
 M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Ulterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: A75289
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-315 <WHI>
 A:Cross-references: GB:AE002063; GB:AE000513; NID:96460121; PIDN:AAFL1861.1; PID:9646
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2317
 A:Map position: 1
 C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

Query Match 22.2%; Score 384; DB 2; Length 315;
 Best Local Similarity 32.4%; Pred. No. 8.2e-23;
 Matches 103; Conservative 55; Mismatches 146; Indels 14; Gaps 8;

Oy 10 RLGRSALLTSRLMIGTVNFSGRVEDDALARLMDHAR---DRCINCLDPTADMGWRLYKGI 66

```

Db      5 KIGKSGILKSEVALGGMETYG--VNQDASKMVGDIYAAAYDEGVNFDQADVYA----RCQ 59
QY      67 TEELVGRMLAOGGGRREDYVLTATKVGSGSERNDGSLARHIIASCEGSLRLGVHDID 126
Db      60 SEQLMKAVLRD--FPHHTLVLSKVFMPMSDDVNDGSLRKHLIESDLSRLGLGDIYD 117
QY      127 VYQMHHIDRSPWDEWQAMDSLVASGVSYSSNFAGMHTAAOENARHSLGMVSH 186
Db      118 IYFAHRYDEVPMEETIVMAFDQYIRDGKALYMGTSWMPARIAQAVEPAKANGLHAPVE 177
QY      187 QCLYINLAVR-HAEVLVPAQAQVGLGVFAPASPLHGLISGAL-EKLAAGTAVKSAQGRA- 243
Db      178 QPEYSVWRDRDVEQELLPYTEGAGTGLVWSPPLAMGLTGKRYDEGRPEGARLTKEKNMAG 237
QY      244 QVLLPRLPAIEAEYKFCNKLGEDPAEVLGAWLVSRRPGIAGAVIGRTPEDLSALKASA 303
Db      238 SYLTEDNICKVNDKSLIADDLGTRAQALAMLLRQKGVSVITGATKYNQIDYTKAAG 297
QY      304 WRLDEQALSELDEIF-PA 320
Db      298 VRLSDVQRRIEDILKPA 315

```

RESULT 13

```

T52133
potassium channel beta subunit homolog [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000
C:Accession: T52133
R:Spoomak, P.; Palme, K.
submitted to the EMBL Data Library, April 1998
A:Description: Potassium channel beta subunit homolog.
A:Reference number: 225974
A:Accession: T52133
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-328 <SPO>
A:Cross-References: EMBL:AF061570; PIDN:AAC15999.1
C:Genetics:
A:Gene: KV-beta1
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

```

```

Query Match      22.0%; Score 380.5; DB 2; Length 328;
Best Local Similarity 30.6%; Pred. No. 1.6e-22;
Matches 102; Conservative 64; Mismatches 144; Indels 23; Gaps 9;

```

```

QY      11 LGRSALLTSLRLMIGT-VNFSGRVEDDALARLMDHARDGINCIDTADMTGWRLYKGHTTE 69
Db      6 LKSGIKVSTLSFGAWVTFCGNQIDVKEAKSILOCCRDHGVNFDMAEYVA---NGRAEE 61
QY      70 LVGRMLAOGGGRREDYVLTATKY--GGEMSERVNDGSLARHIIASCEGSLRLGVHDID 127
Db      62 IMGQAIRLGLWRRSDIVITKIFWGG---GPPNDKGLSKRHIVETGKASKSLKMDMYVDV 118
QY      128 YQMHHIDRSPWDEWQAMDSLVASGVSYSSNFAGMHTAAOENARHSLGMVSH 187
Db      119 LVCHRDASTPIETRYRANVYIDKGMAYFWGTSEMSAQOITFAMWCAADRDLVGIVQ 178
QY      188 CLYINLAVR-HAEVLVPAQAQVGLGVFAPASPLHGLISGAL-EKLAAGTAVKSAQGRA- 242
Db      179 PEYNMFAKKRKEVEEFLPLTNHIGITLWTWSPPLASGVLTGKYNKGAIPDSRFALENVKML 238
QY      243 A-QVLLPRLPAIEAEYKFCNKLGEDPAEVLGAWLVSRRPGIAGAVIGRTPEDLSALKASA 301
Db      239 ANBSLVDVLRKVGSLKPIADELGVTLAQDALIWCASNPVSVITGATRESQIDENMKA 298
QY      302 SAMTDEQALSE---LDEIFPAVASGGAPEAW 331
Db      299 ----VDVILPLPIVLDKIEQVYIGSKFRPEST 327

```

```

RESULT 14
B83093
probable oxidoreductase PA4434 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83093
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MOID:20437337
A:Accession: B83093
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-345 <STO>
A:Cross-References: GB:AE004857; GB:AE004091; NID:g9950654; PIDN:AA007822.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4434
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

```

```

Query Match      21.6%; Score 373; DB 2; Length 345;
Best Local Similarity 32.0%; Pred. No. 6.7e-22;
Matches 108; Conservative 54; Mismatches 138; Indels 38; Gaps 10;

```

```

QY      11 LGRSALLTSLRLMIGT-VNFSGRVEDDALARLMDHARDGINCIDTADMTGWRLYKGH 66
Db      6 LGRITDKVSAKLCIGTWTWQNSQDAFAOIAKAKAGINFMDTAEYVPPAEYV-AS 64
QY      67 TEELVGRMLAOGGGRREDYVLTATKVGSGS--ERVNDGSL--SARHIIASCEGSLRLGV 122
Db      65 TERIIQNMFRSGD--RADWILASKIPGNGISHVDGMLKFNROHIVALDLSERLQT 123
QY      123 DHIDYQMH-----HIDRS-APWDEWQAMDSLVASGVSYSSNFAGM 166
Db      124 DWLDLYQLHMPERRINPFGQLGYQHGESEFTPLETLEVLDEGVRAKIRHIGLSMETW 183
QY      167 HIAAOENARHSLGCVSHOCLYNLAVRHAEVLVPAQAQVGLGVFAPASPLHGLISGA 226
Db      184 GMTFTFLRAEERKMPRAVSIQNPYNLINSFVGLAEIARRECGGLATSPNAFGMLSG- 242
QY      227 LKRLAAGTAVKSAQGAQVLL-----PSLRPAIEAEYKFCNKLGEDPAEVLGAWLVS 279
Db      243 --KYADG--ARPANRISLVSRTRYTNPOAEAKARVVALAREHGLEPQMALAVYTSR 298
QY      280 PGIAGAVIGRTPEDLSALKASAMTLDQALSELDEI 317
Db      299 PFTVSNIGATSLDELETNIGSVDLRLDEEVLAGIDAI 336

```

RESULT 15

```

S61978
hypothetical protein YPL088w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein Lpg20w
C:Species: Saccharomyces cerevisiae
C:Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 18-Feb-2000
C:Accession: S61978
R:Yang, Y.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Hall, J.; Storms, R.K.;
submitted to the EMBL Data Library, December 1995
A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A:Reference number: S61959
A:Accession: S61978
A:Molecule type: DNA
A:Residues: 1-342 <MAN>
A:Cross-References: EMBL:U43281; NID:g1151218; PIDN:AA068211.1; PID:g1151238; GSPDB:G
C:Genetics:
A:Gene: MIPS:YPL089c
A:Map position: 16L
C:Superfamily: fission yeast pyridoxine 4-dehydrogenase

```

	Query Match	21.5%	Score	371.5;	DB 2:	Length	342;			
	Best Local Similarity	28.9%;	Prod No.	8.7e-22;						
	Matches	97;	Conservative	67;	Mismatches	135;	Indels	37;	Gaps	8;
Oy	9	VRLGRSALLTSRLMLGTGVNFSGR-----VEDDQAL-RIMDHARDRGINCILDTADMYGMR	61	.						
Dd	7	VRLNGSLGLISPIVIGCMSYGSKKWADWDIEDKXQIFPKMKCYDKGLRPTTADFY--	64							
Oy	62	LYKGHTEELVGMWLAAOGGRRDDYLAIRKVGGEMSERV-----NDSGI	104							
Dd	65	--NGLSERLIKFELEYYSIKRETVIMTKIYPVDETLDLHHNFTLNFEEDLDSNQRL	122							
Oy	105	SARHTIASCEGSLRLRVGHIDVYMHHMDRSAPDEWQAOMDSLVAAGSVYGSSMNA	164							
Dd	123	SKKHIIAGVENSVKRIQT-YIDLDOIHRDLHETPMKEIKALKINDVVEGAHWARTGASSKL	181							
Oy	165	GWHIAAOENARRHSRLGNVSHOCYTNLAVRAAELEVPAAOAYGLGFAMSPLEGGLTS	224							
Dd	182	APEFELQPTADKYMPQFISSQYYNNLLYREDRELIPFARNHNIGLLPMSPPNRMGMT	241							
Oy	225	GALEKLAGTAIVKSAQGARAQVILLPSL---RPATAYEKFCCNLGEDPAENGLAVLSRP	280							
Dd	242	RPLNQ--STDRIKSDPTFKSLHMDLMEEBOKELINRREVKSODKRYVSAMLSIAVYLHG	299							
Oy	281	GIAGAVIGPRPEQLDSALKASAMTLDOQASELNE	316							
Dd	300	--CHPIVGINTARVDEAIALAQVLTTFEEKITLEE	333							

```
Search completed: June 22, 2001, 09:25:12
Job time: 107 sec
```

W

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: June 23, 2001, 10:26:43 ; Search time 113.8 Seconds

5512.071 Million cell updates/sec

Title: US-09-463-705A-1_COPY_48_1046

Sequence: 1 ctgcaaccagcttcgcgcg.....gtcgcgcgtcgtgtgtcat 999

Scoring table: IDENTITY_NUC

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_0601:*

1:	/SIDSR_gcgdata/genseq/genseqn/NA1980.DAT *
2:	/SIDSR_gcgdata/genseq/genseqn/NA1981.DAT *
3:	/SIDSR_gcgdata/genseq/genseqn/NA1982.DAT *
4:	/SIDSR_gcgdata/genseq/genseqn/NA1983.DAT *
5:	/SIDSR_gcgdata/genseq/genseqn/NA1984.DAT *
6:	/SIDSR_gcgdata/genseq/genseqn/NA1985.DAT *
7:	/SIDSR_gcgdata/genseq/genseqn/NA1986.DAT *
8:	/SIDSR_gcgdata/genseq/genseqn/NA1987.DAT *
9:	/SIDSR_gcgdata/genseq/genseqn/NA1988.DAT *
10:	/SIDSR_gcgdata/genseq/genseqn/NA1989.DAT *
11:	/SIDSR_gcgdata/genseq/genseqn/NA1990.DAT *
12:	/SIDSR_gcgdata/genseq/genseqn/NA1991.DAT *
13:	/SIDSR_gcgdata/genseq/genseqn/NA1992.DAT *
14:	/SIDSR_gcgdata/genseq/genseqn/NA1993.DAT *
15:	/SIDSR_gcgdata/genseq/genseqn/NA1994.DAT *
16:	/SIDSR_gcgdata/genseq/genseqn/NA1995.DAT *
17:	/SIDSR_gcgdata/genseq/genseqn/NA1996.DAT *
18:	/SIDSR_gcgdata/genseq/genseqn/NA1997.DAT *
19:	/SIDSR_gcgdata/genseq/genseqn/NA1998.DAT *
20:	/SIDSR_gcgdata/genseq/genseqn/NA1999.DAT *
21:	/SIDSR_gcgdata/genseq/genseqn/NA2000.DAT *
22:	/SIDSR_gcgdata/genseq/genseqn/NA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	999	100.0	3412	20	AAK25772	S. erythraea eryth
2	999	100.0	3756	18	AAV72684	Sugar biosynthesis
3	511.6	51.2	9994	22	AAC85191	S. avermitilis 10
4	150	15.0	1512	22	AAFE61103	p. putida KT2440-e
5	128.2	12.8	1811	21	AAAC44737	z. mays DNA fragm
6	92	9.2	1032	19	AAV50486	Streptomyces clav
7	92	9.2	7193	19	AAV50431	Streptomyces clav
8	84.6	8.5	4257	19	AAV68520	The nucleotide seq
9	84.6	8.5	4257	19	AAV10362	Infected cell prot
10	83.8	8.4	114955	20	AAK53491	Human adenosine A1
11	81.8	8.2	114955	20	AAK53491	Human adenosine A1

C 12	79.8	8.1	3849	22	AAE257559
C 13	81.2	8.0	58857	21	AAE584711
C 14	77.8	7.8	44377	18	AAE785088
C 15	77.4	7.8	44377	18	AAE804154
C 16	77.7	7.7	15872	18	AAE867155
C 17	75.8	7.6	12001	16	AAE076213
C 18	75.8	7.6	15872	21	AAE877283
C 19	75.8	7.6	43280	18	AAE04133
C 20	75.2	7.5	1896	21	AAE849637
C 21	73	7.3	3957	22	AAE096666
C 22	72.8	7.3	3978	21	AAE557804
C 23	72.8	7.3	12249	21	AAE558406
C 24	72.8	7.3	18331	21	AAE558575
C 25	72.6	7.3	37856	21	AAE119927
C 26	72.2	7.2	2061	10	AAAN2408
C 27	71.8	7.2	2943	17	AAE164808
C 28	71.8	7.1	1998	12	AAE011277
C 29	71.2	7.1	5392	15	AAE042011
C 30	71.2	7.1	2776	18	AAE873070
C 31	71.1	7.1	30001	18	AAE101016
C 32	71.1	7.1	30001	20	AAE05110
C 33	70.7	7.1	11220	21	AAE287288
C 34	70.6	7.1	13613	21	AAE27319
C 35	70.6	7.1	36778	21	AAE273738
C 36	70.6	7.1	37948	21	AAE275855
C 37	70.7	7.1	38506	21	AAE56533
C 38	70.6	7.1	38506	21	AAE56001
C 39	70	7.0	1044	20	AAE23162
C 40	70	7.0	4390	20	AAE231820
C 41	69	6.9	2918	21	AAE243662
C 42	69	6.9	2919	21	AAE243922
C 43	67.8	6.8	23666	12	AAE019100
C 44	67.4	6.7	671	12	AAE12227
C 45	67	6.7	833	15	AAE042033

ALIGNMENTS

post-date

ID	AAAX25772	standard; cDNA; 3412 BP.
AC	AAAX25772;	
DT	08-JUN-1999	(first entry)
XX		
DE	S.erythraea erythromycin-synthesis gene cluster eryG-eryAIII.	
XX		
KW	gene cluster; bacterium; enzyme; macrolide; antibiotic; erythromycin;	
KW	secondary metabolite; eryBII; eryCII; eryCIII; hybridisation; probe;	
KW	glycosylation; macrolactone; oleandomycin; ds.	
XX		
OS	Saccharopolyspora erythraea.	
XX		
Key	Location/Qualifiers	
EH	complement (47..1048)	
FT	/*tag= a	
FT	/label= ORF7	
FT	/gene= "eryBII"	
FT	/product= "dTPD-4-keto-L-6-deoxyhexose-2,3-reductase"	
FT	complement (1045..2310)	
FT	/*tag= b	
FT	/label= ORF8	
FT	/gene= "eryCIII"	
FT	/product= "desosaminyltransferase"	
FT	complement (2321..3406)	
FT	/*tag= c	
FT	/label= ORF9	
FT	/gene= "eryCII"	
FT	/product= "dTPD-4-keto-D-6-deoxyhexose-3,4-isomerase"	
XX		
NN		

post doc

FT	CDS	complement (6451..7845)
FT	FT	/tag= g
FT	FT	/product= "oleandrose synthase"
FT	FT	/gene= "AvrG"
FT	FT	/note= "ORF6 protein AAB61979"
FT	CDS	complement (7858..8631)
FT	FT	/tag= h
FT	FT	/product= "Glycosyl methyltransferase"
FT	FT	/gene= "AvrH"
FT	FT	/note= "ORF7 protein AAB61980"
FT	CDS	complement (8718..9761)
FT	FT	/tag= i
FT	FT	/product= "oleandrose synthase"
FT	FT	/gene= "AvrI"
FT	FT	/note= "ORF8 protein AAB61981"
XX	XX	W0200109155-A1.
XX	XX	08-FEB-2001.
XX	XX	26-JUL-2000; 2000MO-US20331.
XX	XX	30-JUL-1999; 99US-0146699.
XX	XX	(MERI) MERCK & CO INC.
XX	PI	MacNeil DJ, Occi J, Gewain KM;
XX	DR	WPI; 2001-182929/18.
XX	DR	P-PSDB; AAB61973, AAB61974, AAB61975, AAB61976, AAB61977, AAB61978.
XX	XX	Novel nucleic acid fragments of Streptomyces avermectilis genome useful
XX	XX	for enzymatic, biochemical, biosynthetic and diagnostic purposes -
XX	XX	Claim 1; Page 5-16; 63pp; English.
XX	XX	The invention relates to a 10 kb genomic DNA isolated from <i>S. avermectilis</i>
XX	XX	that contains genes encoding proteins for glycosylation of avermectin
XX	XX	aglycones. The polypeptides can be expressed by standard recombinant
XX	XX	methodology. The gene and the encoded polypeptides together with other
XX	XX	enzymatically active polypeptides, are useful to perform combinatorial
XX	XX	biocatalysis in vitro and in vivo in a host cell. They are useful for
XX	XX	performing biotransformations on macrocyclic compounds including avermectin
XX	XX	or other macrocyclic aglycones. The sequences are also useful in vivo in a
XX	XX	bacterial host, in vitro in combination with an actinomycete
XX	XX	fermentation, and in vitro in combination with enzymatically active
XX	XX	polypeptides that are not from the avermectin biosynthetic pathway to
XX	XX	effect the synthesis of a pharmaceutically active compound, e.g. an
XX	XX	antibiotic. The present sequence represents the 10 kb genomic fragment
XX	XX	that contain genes involved in the synthesis and/or addition of
XX	XX	oleandrose to avermectin aglycones.
XX	XX	Sequence 9994 BP; 1461 A; 3204 C; 3705 G; 1624 T; 0 other;

[illegible]

Dd	8904	gcgacaaatcgtggcggtgcctgaagcgccgtggttcctcagcaacaagtctctctgttaacgtccagatggtt	8695
Oy	244	cgggcgcagggagacggccaagaaccctctgcagacggcccttgccgcgcgaacttcaacgcgggtgcc	303
Dd	8964	ggctcagctgtgtgtcatccaaaggacctctcaaacgcgcgcgcctctgtccaggaattgaacgcggtgtt	9023
Oy	304	cggcgagcaagctctcccaagcgcctcgcctctgaagcaaggccgcgcgtgtcaagcgggacaaagcgaaa	363
Dd	9024	ctcccgcgtcctctgcgcgaaggacaacgcgcgcgcgaagcgcgcgcgtcgaatgaatggcgcgaacaacag	9083
Oy	364	gaacgcgaacgcgcgttagagccccgc	423
Dd	9084	gaacgcgaacgcgcgttagagccccgc	9143
Oy	424	caagttcttaacagcacctgtgtgaggagaacatactgcacaggaagtgagcggcgcgcggtctctc	483
Dd	9144	gaagttcttaacagcacctgtgtgaggagaacatactgcacaggaagtgagcggcgcgcggtctctc	9203
Oy	484	ctgc	543
Dd	9204	ctgc	9263
Oy	544	gctgc	603
Dd	9264	ctgc	9323
Oy	604	gtgtgtcatctgtgtgaagcgttcgaatgtgtgttcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgc	663
Dd	9324	gtgtgtcatctgtgtgaagcgttcgaatgtgtgtgttcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgc	9383
Oy	664	ggagagcgatgatgtgtgc	723
Dd	9384	ggagagcgatgatgtgtgc	9443
Oy	724	gcccgaaccttgtctgc	783
Dd	9444	gcccgaaccttgtctgc	9503
Oy	784	gcccacaagctctcctgtgtgtgcgcctctgtgaagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	843
Dd	9504	gcccgaacctctcctgtgtgtgtgcgcctctgtgtgaagacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	9563
Oy	844	gaagcgatgatgtgc	903
Dd	9564	gaacagagtgatgtgc	9623
Oy	904	gaacgc	963
Dd	9624	caaccgc	9683
Oy	964	cccgag 969	
Dd	9684	gcccaag 9689	
RESULT 4			
ID	AAF61103/c		
XX	AAF61103 standard; DNA; 1512 BP.		
AC	AAF61103;		
XX	16-MAY-2001 (first entry)		
DE	P. putida KT2440-associated DNA ORF11654.		
XX			
KM	Transgenic plant; detection; probe; amplification; vaccine carrier;		
XX	Microbial production strain; biological remediation; ds.		
OS	Pseudomonas putida.		
XX	DEL19935088-A1.		
PN			
DD	01-FEB-2001		

RESULT	4
AAFe1103/c	
ID	AAFe1103 standard; DNA; 1512 BP.
XX	
AC	AAFe1103;
XX	
DT	16-MAY-2001 (first entry)
XX	
DE	P. putida KT2440-associated DNA ORF11654.
XX	
KW	Transgenic plant; detection; probe; amplification; vaccine carrier
XX	microbial production strain; biological remediation; ds.
OS	<i>Pseudomonas putida</i> .
XX	
FN	DE19935086-A1.
XX	
PD	01-FEB-2001.

PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 18-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139452.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139753.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142290.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.

PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147935.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148319.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148341.
PR 16-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 22-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151303.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155559.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.

OY	207	cgaggttgcgagcaaaccttctcgaagccctcgtatcgccgagagccg-----aggagag	255
Db	814	CCAGACCGCTCCGCCACCGCGGCGCAGCCGGGTGCACCAACGAGCGATTGGCGCGCAGGGGCGG	755
OY	258	gcaaaagaaacacctgtgcagacagcgccctgtgcgcaacttcaacgcggtgtcccgcggaacttct	317
Db	754	GAGCGCTGAACCGGGGGCTGGGCCCCGGGGAAGTCCCTGTCCTCCCAAGATGCTCGGTGTGC	695
OY	318	ccagagcttcgcgtctgaagcagcgccgctgtcagcagcgagcaacgcggaagaacgcagccgct	377
Db	694	GGATGTGTCCCGCTGATGAGAAACCCCGTCCAGAGGGGGGTAAAGGAGCATCCGATCCCA	635
OY	378	agagccttgcgagcggaacgaacccctccagcttcctgcgtgtcgcgaacgcagctgtttaagc	437
Db	654	GCTCCCGGACAGAGGGACACCACTCTGTCTGTGATCCGCGCGACACAGGCTTCACATTCG	575
OY	438	actgtgtgagaaacatgtcccaaggagttgagcgcgagcgcggttctcctgtgcgagcgga	497
Db	574	TCTG-----CACGGCGGTACCGGGGTGCAACCGGTCTCCGCGGGCGAGGTGTGGCGGG	521
OY	498	tgtgcagacccgcggaagttcgaacgacgcgaagttagaagaccttgcgctgtgcgaagagc	557
Db	520	AGGGCTCG-----GAGAGACCGAGCCTGCGGACCTTGCCCTCGCGCACAGACT	473
OY	558	tgtcatagccttcgcacacactcgtcccaacgagcgcgagccgcttgatgtgtgtcatgtgt	617
Db	472	CGGCCACCGCACCCACGCTCTCTCTGATCGGACACCGCGGGTCCGTCACTGTGTGTA	413
OY	618	agaagctcgaatgtgtgtcagacgcccacgcttgcagcagcagatcccttcgcagagagcgatagt	677
Db	412	ACAAGTGCATCGGTGCGGTGCGGACGAGACGAGGACCGTTCGCAAGGCGCGGAGCT	353
OY	678	gcgcgcgcgaagacccgctgtgtgtacgcgctcgtctcatcttcgcgcgcgaactgtgtcg	737
Db	352	AGGACGGC-----TCGCGGACAAAGCCTTGGAAGGCGCGCTCGACGAGAC	308
OY	738	ccagacaggtgtctctcgcgcgcttcgcgcgccttgcagcacaacccgcccacagctct	797
Db	307	GCACCATCCCAATTGTTGTGGCATACGACACTGTCCGGCGGGCCGACCGCCCGTC	248
OY	798	cgaatgtgtccctgtatagaccgcgcgaacgcgtacatgttcgcgctgttcgagagcagttatgc	857
Db	247	CGACGAGCTCTCCACCGCGCGCCGACCCCTCGAGCGTCCGCGGTTCTCAGAGGAGTGACC	188
OY	858	cgagctcccgaggtgtgtcatcaagcgacgcgcgctgcgtctcctcgaagcgttcgctga	917
Db	187	CGGCGTGCAGCGCGCGCGCGAGTGTGGCTGCTCCCGGGCGGTCCTCGGCGCTCTCTGGA	128
OY	918	agttaccagctgcgcagcagcag	937
Db	127	AGTCGATGTGTGCGCAGGCGAG	108
RESULT	7		
AAV50431			
ID	AAV50431	standard; DNA; 7193 BP.	
XX	AAV50431:		
AC	26-OCT-1998	(first entry)	
XX			
DE		Streptomyces clavuligerus DNA sequence of ORFs flanking casI.	
XX			
KW		Streptomyces clavuligerus; bacterial gene; clavulanic acid; 5S clavum	
KM	5R	clavum; CASI; ORF; open reading frame; biosynthesis; ss.	
XX			
OS		Streptomyces clavuligerus.	
XX			
XX	MO9833896-A2.		
XX			
TD	06-AUG-1998.		
XX			

PF 02-FEB-1998; 98WO-EP00644.
 PR 04-FEB-1997; 97GB-0002218.
 XX (SMIK) SMITHKLINE BEECHAM PLC.
 PA (UYAL-) UNIV ALBERTA.
 XX
 PI Anders C, Barton B, Griffin JP, Jensen S, Mosher RH;
 PI Parakkhar AS;
 XX
 DR WPI: 1998-437451/37.
 XX
 PT DNA comprising defective 5S clavin biosynthesis gene(s) from
 PT Streptomyces clavam - useful for producing clavulanic acid without
 PT production of 5S clavam or clavam-2-carboxylates
 PS
 PS Claim 2; Page 15-24; 29pp; English.
 XX
 CC The present sequence represents DNA comprising genes specific for 5S
 CC clavam biosynthesis from Streptomyces clavuligenus (SC) and which are
 CC not essential for 5S clavam biosynthesis. The present invention also
 CC describes: (1) a process for improving 5R clavam production in a
 CC suitable microorganism comprising manipulation of DNA as above and its
 CC inclusion in the microorganism; (2) a process for improving 5R clavam
 CC production in SC comprising disrupting or otherwise making defective DNA
 CC regions flanking cas I; (3) a process for the identification of a
 CC microorganism suitable for high 5R clavam production comprising a
 CC preliminary screening for microorganisms with low or no 5S clavam
 CC production; (4) a microorganism which is capable of 5R clavam production
 CC and low or no 5S clavam production obtainable by a process as in (3);
 CC (5) clavulanic acid (CA) obtainable by the fermentation of a
 CC microorganism as in (4), and (6) CA which is free of any 5S clavam; (7)
 CC CA which is free of any clavam-2-carboxylate. The methods and products
 CC can be used to produce organisms capable of producing increased amounts
 CC of clavams suitably e.g. CA, for use as antibiotics. The methods can
 CC also be used for the production of CA without the production of 5S
 CC clavam or clavam-2-carboxylate.
 XX
 XX Sequence 7193 BP; 1055 A; 2745 C; 2498 G; 895 T; 0 other;

[illegible]

[illegible]

XX	RESULT	8	
XX	AAV68520/c		
XX	ID	AAV68520	standard; DNA; 4257 BP.
XX	AC	AAV68520;	
XX	DT	29-JAN-1999	(first entry)
XX	DE	The nucleotide sequence of the Herpes simplex virus ICP4.	
XX	XX		
XX	XX	Herpes simplex virus ICP4; HSV ICP4; modulator; apoptosis;	
XX	KW	stimulation; inhibition; HSV infection; ss.	
XX	OS	Herpes simplex virus.	
XX	XX		
XX	FT	Key	Location/Qualifiers
XX	FT	CDS	361..4257
XX	FT		/*tag= a
XX	FT		/product= "HSV ICP4 protein"
XX	XX		
XX	PN	WO9846637-A2.	
XX	PD	22-OCT-1998.	
XX	XX		
XX	PF	16-APR-1998;	98WO-US07573.
XX	PR	16-APR-1997;	97US-0843659.
XX	PA	(ARCH-) ARCH DEV CORP.	
XX	PI	Leopardi R, Rolzman B;	
XX	DR	WPI: 1998-594559/50.	
XX	DR	P-PSDB: AAW80810.	
XX	XX		
XX	XX	Use of herpes simplex virus (s)3 polypeptide - for developing	

PT products for modulating apoptosis in cells and for identifying
PT compounds which act as stimulators or inhibitors of apoptosis
XX
PS Example 2; Pages 60-63; 85pp; English.

CC This is the nucleotide sequence of Herpes simplex virus ICP4 used in
CC the method of the invention as modulators of apoptosis. The methods
CC and products can be used to identify compounds which modulate
CC (stimulate or inhibit) apoptosis in cells. They can be used to
CC immortalise cells for the study of these cells or for growing cells
CC in large numbers for the productions of proteins. They can also be
CC used for stimulating apoptosis in cells, e.g. for treating a subject
CC with a HSV infection.

Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T; 0 other;

Query Match	8.5%;	Score 84.6;	DB 19;	Length 4257;
Best Local Similarity	42.9%;	Pred. No. 2.2e-05;		
Matches 423; Conservative	0;	Mismatches 564;	Indels 0;	Gaps 0;

Dy 10 ggtcttcgagcgccgcgcgcgcgaagcaccgcgggagaagtctcgttcaatttcyaacg 69
||| ||| | | | | | | | | | | | |
Db 2490 GGGCTCCGGTGCCTGGCGGCCGCGCAGACGAGGGGCGCAGGCTTGTCACAACAG 2431

DY 70 cgcctcgtccagtgatcgacgaccttcacgcgcgaagtctgactgccyygggt 129
| | | | |
Db 2430 CAGGTCCCGCGGCGCGCGCGGAGCTGGCAGGC GCGGTCGCCGCGGCAACGCGCGG 2371

Dy 130 tctgcggtccgatgaacgcgcgcgcgatgccggcagcaggcaaccatctgcgccccaac 189
| | | | | | | | | | | | | | |
Db 2370 gccacggcccccgccgaccaggctcacggcgccacaggcgccacggcgcccttcgtgccc 2311

Db 2310 GCCGGCCACCGCAGTCCCCGCCGACGCCTGAGCACCAAGCGCGTCGCCGACGAAACC 2251

Db 2250. CAGCTCGCGCACGCCGCGCAGCGCGGGCGGTGCGCGTTCGCGCGCGCGCGGGGAGC 2192

D_b 2190 GGGGCCCCGCCCTCCGGGCCGGGGGGCGTCGGGGGCCGGCCCCCGGCAC6CCCCGG 2131

Db 2130 GACGGCCGCCAGTTCGCCCTCGAAGCCCTCGGGCAGCGGCTCCAGATCCCGGGCAGGC 2071

2070 GGGCAGGCACCTGACAGGCCACAGCGGCGGGCTGGGGCGGCGCGCCCGGCGTCTGTCGTCGGC 2011

2010 GTGGCGTGGCGGCGGTCTGCGCCCCCGGGGAGGCGCGCGCGGA 1951

Db 1950 CAGCCGCCCCAGGGCGGGCGAGGATCCCCCGGGCGCGGTACCCGGGGCACCAGCGCGTTC 1991

Db 1890 GCGCGTGGGGGGGGGACGGGGGACCCCTGTCATCTGGCGGGGGGGGGT 1831

Db 1830 CCCCCGCGCGTACGCGCGGCTTCGCGCGCCACAGGGGCGCGTAGCGCGGCG 1771

Db 1770 CAGGCTGGTCACGACGAAAGCCCTTCTGCGCGGCTGATCGCGGCTCATGCGCCACAGGC 1711

Accession	Sequence	Position
OY	caagctctcgtgtgtgcccccttctgtaagagccgcagacgcagatcatatctcggcggtctcgagca	849
Db	1710 gggccggccgctgtgcggccacagagcccccacggcgaagcggccggccgacatgggttaccoccatg	1651
OY	850 gtltatgtccgcggtgtcccgagcggtgtgtcattcatgcagcgacagcgcgctgtcgtctcgcagcg	909
Db	1650 gggcagacggcccgccggccacacgctgtgcgggtgattgaagagctcgtgttgccggcgccggccga	1591
OY	910 tcgcgtgaagttaacgcgttcgcgaagccagagccctgtctgtgtgacagcgcggaagccgcag	969
Db	1590 gatccggagacagagccctcgtgtgtccagccccacgctcccgggagacacagccggggttctgagc	1531
OY	970 ccgcagctgtcgtcggcgagcgtcagttgt	996
Db	1530 ccaccccatgagccctccgcgcgtccggggt	1504

RESULT	9
AAV10362/c	
ID	AAV10362 standard; cDNA; 4257 BP.

AC AAV10362;

DT 30-JUN-1998 (first entry)

DE Infected cell protein number 4 alpha-4 gene.

KM Infected cell protein number 4; ICP4; alpha-4; cell apoptosis;
KW therapeutics; ss.
..

Herpes simplex virus.

FH	Key	Location/Qualifiers
FE	cond	361 4357

```

FT      /product= "Infected cell protein"

```

PN WO9804709-A2.

PD 05-FEB-1998.

PF 23-JUL-1997; 97WO-US12904.

PR 26-JUL-1996; 96US-0690473.

PA (ARCH-) ARCH DEV CORP.

PI Leopardi R, Roizman B;

DR WPI; 1998-130697/12.

XX
DE

PT blocking apoptosis in cells, production of proteins and gene therapy

PS Disclosure; Fig 2; 63pp; English.

CC The herpes simplex virus (HSV) alpha-4 gene encodes infected cell protein
CC number 4 (ICP4). Both the alpha-4 gene and ICP4 can be used to block
CC cell apoptosis. Similarly the administration of an agent that inhibits
CC ICP4 or the alpha-4 gene can induce apoptosis in HSV infected cells.
CC This can be used for the immortalisation of cells, production of
CC proteins, gene therapy, or inhibition of cell death induced in vivo.
CC They can also be used for production of therapeutics comprising
CC inhibitors of HSV ICP4 function, useful for treating HSV function.
XX Sequence 4257 BP; 412 A; 1768 C; 1663 G; 414 T; 0 other;

Query Match	8.5%;	Score 84.6;	DB 19;	Length 4257;
Best Local Similarity	42.9%;	Pred. No. 2.2e-05;		
Matches 423;	Conservative	0;	Mismatches 564;	Indels 0;
				Gaps 0;

[illegible]

RESULT	10
AA53491	
ID	AA53491 standard; DNA; 114955 BP.
XX	

PE 16-JUN-1999; 99DE-1028313.
 XX
 PR 16-JUN-1999; 99DE-1028313.
 XX
 PA (KELL/) KELLER U.
 XX
 PI Schawecker F, Keller U;
 XX
 XX
 DR WPI: 2001-081744/10.
 XX
 PT Introducing N-methyltransferase activity into peptide synthetase
 PT enzymes, useful for synthesis of N-methylated peptides such as
 XX antibiotics, by altering DNA -
 XX
 PS Example 2; Page 9; 18pp; German.
 XX

CC This invention describes a novel method of introducing an
CC N-methyltransferase domain (A) into peptide synthetase (PPS) activation
CC domains by altering the DNA that encodes the domains. The invention also
CC describes a method for combining (or gene segments) that encode PPS
CC modules that lack (A) with genes (or segments) encoding modules that
CC include (A). Modified DNA sequences formed by insertion of (A) are used
CC (1) for altering natural (or already altered) PPS or polypeptide
CC synthetase (PPS) genes and their fragments, also for construction of new
CC PPS and PPS gene and (ii) for construction of plasmids or genetically
CC altered organisms for synthesis of encoded proteins (B). (B) are used for
CC in vivo or in vitro enzymatic syntheses of amino acids, polypeptides, and
CC peptidyl-acetyl mixed structures containing N-methylated amino acids, or
CC their derivatives, also for fermentative production of such compounds.
CC These compounds are often of pharmaceutical value, e.g. penicillins,
CC vancomycin, cephalosporin, pristinamycin or actinomycin D. The modified
CC (B) enzymes involved in non-ribosomal peptide synthesis are able to
CC N-methylate their substrates but substrate specificity remains unchanged.
XX
SQ Sequence 3849 BP; 596 A; 1605 C; 1122 G; 446 T; 0 other:

Query Match	8.1%;	Score 81.2;	DB 22;	Length 3849;
Best Local Similarity	44.0%;	Pred. No. 7.4e-05;		
Matches 399;	Conservative 0;	Mismatches 498;	Indels 9;	Gaps 1;

[illegible][illegible]

RESULT	13
AAA58471/c	
ID	AAA58471 standard; DNA; 58857 BP

AC	AAA58471;
XX	
DT	31-OCT-2000 (first entry)
vv	

DE	Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30
XX	
KW	BIM gene cluster; bleomycin gene cluster; polyketide metabolites;
KW	bleomycin; bleomycin analogue; holo carrier protein; thiazolidine;
KW	thiazoline; dithiazoline; microbial metabolite; sugar; ss.
XX	
OS	Streptomyces verticillus.

FT	Key	Location/Qualifiers
FT	CDS	223..564
FT		/*tag= a
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 30; encodes AAB07556"
FT	CDS	561..2309
FT		/*tag= b
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 29; encodes AAB07557"
FT	CDS	2767..3486
FT		/*tag= c
FT		/note= "ORF 28; encodes AAB07558"
FT		3527..5593
FT	CDS	/*tag= d
FT		/transl_except= (pos: 1..3, aa: Met)
FT		/note= "ORF 27; encodes AAB07559"
FT		5806..12294
FT	CDS	/*tag= e
FT		/note= "ORF 26; encodes AAB07560"
FT		12291..15491
FT	CDS	/*tag= f
FT		/note= "ORF 25; encodes AAB07561"
FT		15488..21013
FT	CDS	/*tag= g

FT	/note= "ORF 24; encodes AAB07562"
FT	21010..24666
FT	/*tag= h
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 23; encodes AAB07563"
FT	24663..32690
FT	/*tag= i
FT	/note= "ORF 22; encodes AAB07564"
FT	32893..34830
FT	/*tag= j
FT	/note= "ORF 21; encodes AAB07565"
FT	34827..35804
FT	/*tag= k
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 20; encodes AAB07566"
FT	35818..37302
FT	/*tag= l
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 19; encodes AAB07567"
FT	37299..39215
FT	/*tag= m
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 18; encodes AAB07568"
FT	39301..47181
FT	/*tag= n
FT	/note= "ORF 17; encodes AAB07569"
FT	47178..49985
FT	/*tag= o
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 16; encodes AAB07570"
FT	49982..51001
FT	/*tag= p
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 15; encodes AAB07571"
FT	50998..52386
FT	/*tag= q
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 14; encodes AAB07572"
FT	52383..52946
FT	/*tag= r
FT	/note= "ORF 13; encodes AAB07573"
FT	53018..54190
FT	/*tag= s
FT	/note= "ORF 12; encodes AAB07574"
FT	54187..55824
FT	/*tag= t
FT	/note= "ORF 11; encodes AAB07575"
FT	55821..56093
FT	/*tag= u
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 10; encodes AAB07576"
FT	56090..57586
FT	/*tag= v
FT	/transl_except= (pos: 1..3, aa: Met)
FT	/note= "ORF 9; encodes AAB07577"
FT	57583..58857
FT	/*tag= w
FT	/note= "ORF 8; encodes AAB07578"
XX	
PN	W0200040704-A1.
XX	
PD	13-JUL-2000.
XX	
PE	06-JAN-2000; 2000OW-US00445.
XX	
PR	06-JAN-1999; 99US-0115435.
PR	05-FEB-1999; 99US-0118848.
PR	05-JAN-2000; 2000US-0477962.
XX	
PA	(REGC) UNIV CALIFORNIA.
XX	
XX	Shen B, Du L, Sanchez C, Chen M, Edwards DJ;

DR	WP1: 2000-465974/40.
DR	P-PSDB: AAB07556, AAB07557, AAB07558, AAB07559, AAB07560, AAB07561,
DR	AAB07562, AAB07563, AAB07564, AAB07565, AAB07566, AAB07567, AAB07568,
DR	AAB07569, AAB07570, AAB07571, AAB07572, AAB07573, AAB07574, AAB07575,
DR	AAB07576, AAB07577, AAB07578.
XX	
PT	New bleomycin gene cluster components useful for peptide and/or
PT	polyketide metabolites, especially bleomycin, production and for
PT	chemically modifying biological molecules -
XX	
XX	Claim 8; Page 97-136; 162pp; English.
XX	
CC	The present sequence represents the BLM (Bleomycin) gene cluster,
CC	containing open reading frames (ORFs) 8-30. The proteins encoded
CC	by the gene cluster are useful for producing peptides and/or polyketide
CC	metabolites, especially bleomycin or bleomycin analogues. They are
CC	also useful for chemically modifying biological molecules to produce
CC	branched methyl groups, and for coupling amino acids and fatty
CC	acids. They may be reacted with an apo-carrier protein and coenzyme A
CC	to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC	domains can be used individually or collectively to produce
CC	thiazolidine, thiazoline, dithiazoline and dithiazoline-containing
CC	microbial metabolites. The BLM gene cluster may also be used to produce
CC	sugars.
XX	
SQ	Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 other:
	Query Match 8.0%; Score 79.8; DB 21; Length 58857#
	Best Local Similarity 45.3%; Pred. No. 8.5e-05;
	Matches 389; Conservative 0; Mismatches 452; Indels 18; Gaps 2;
OY	73 ctgtcttcacagtggtcatcgcgagcccttcagcgcggaatcgtgactcgggggttccg 132
Db	8649 CTCCTCTGTGGGCGGCGCTTGCGACCGCGGCGCGGCACAGTACCGGTCAGCGTCAg 8590
OY	133 cggcgcgatagaggcgccgcgagtctcgggccggagaacaccattggagcccacctc 192
Db	8589 CTCGCCGGCGACGGCGACGAACGGCGCGGCTCCTCTCCACC CGCGCACACACAGCGC 8530
OY	193 ggcgcggtcttcgcgcgaagttagcgcagaattctcgttagagctcgatcgcgcggcgag 252
Db	8529 CTGGCGCGAOCGGGGGTTGGGAGGCCAGACGGGCTCTGGTCTGGGGGCTCCACCCGGAC 8470
OY	253 ggaacgcacaacagcacctgcgcacgycgccttcgcgcgaattcacacgcggcgtagccgcggcgag 312
Db	8469 GCCCGGATCTTCACCTGCGGCTCGGTGGCGCCGAGGTACTCAGAAGTTCGCTCGGCGG 8410
OY	313 -----cttctcagaagctcgcgttaagcaagcgccgctgtagcgcgagaccaaaggaa 363
Db	8409 CCGCGCGGCGACAAGTTCGCGGTGCGGTAGAACCGCGCGCGCTTCGCGCGGGTAATCTCTGGG 8350
OY	364 gaacgcagaacccgatagactctgcggcgcgagaaaccccaagctcggagctgcgcgaacgcg 423
Db	8349 GAGGAAGCGCTTCGCGGTGAGTTCGCGCGCGCCCCAGTAGCCGTGGGCGACACCGCGCGC 8290
OY	424 caagttltagaacgcaactgttggagaaacctatgcccaagggaagttagcgcgcggtcttc 483
Db	8289 GCCGATTCACAGCTTCGCGGACGACGCCGTGGGGAGACGGGGCGCGCGCTGTGTCAGCAG 8230
OY	484 ctgcgcgcgcgcgatgtgtccadgcgcgcgaagtlctgaagcagccagctlagagaaccttgc 543
Db	8229 GTGCGACGTTGGTGTGGAGATCGCGCGCGCCGATGGGACACGGGCGGTAACGGGGCGTGGC 8170
OY	544 gctgcgcgcagagctgtctcalagcctgcacaacctcgttccaagcgcgcgaaacgyltat 603
Db	8169 GGCGTCGACGAGCTGCGGACGCTCCGACGACCCCTCTTCGCTGGGGCCGCTATCTCTTGAC 8110
OY	604 gtgtgtcatctgttagacgttgtatgtgttgtagacccaacgccttgcgcagcatcccttcca 663
Db	8109 GACCGCGTCCCGGAGGCGCCGGAAGCGCTTCCAAGGCTTTCGCGCCGACCGCTTCTCCGCG 8050
OY	664 ggaagcgatatgttgcgcgcgcgaacagccgcgttcgtttagcgcgctcgtcatcatctgcg 723

[illegible]

RESULT 2
US-08-690-473-1/c

: Sequence 1, Application US/08690473
 : Patent No. 5876923
 :
 : GENERAL INFORMATION:
 : APPLICANT: Leopardi, Rosario
 : APPLICANT: Roizman, Bernard
 : TITLE OF INVENTION: HERPES SIMPLEX VIRUS ICPL4 AS AN
 : TITLE OF INVENTION: INHIBITOR OF APOPTOSIS
 : NUMBER OF SEQUENCES: 2
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSEE: Arnold, White & Durkee
 : STREET: P.O. Box 4433
 : CITY: Houston
 : STATE: Texas
 : COUNTRY: USA
 : ZIP: 77210
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/690,473
 : FILING DATE: 26-JUL-1996
 : CLASSIFICATION: 435
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Highlander, Steven L.
 : REGISTRATION NUMBER: 37,642
 : REFERENCE/DOCKET NUMBER: ARCD:239
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: 512/418-3000
 : TELEFAX: 512/474-7577
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 4257 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 :
 : US-08-690-473-1

[illegible]

[illegible][illegible]

[illegible]

RESULT 8
 US-08-804-227C-1/C
 : Sequence 1, Application US/08804227C
 : Patent No. 5876591
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Dehoff, Bradley S.
 : APPLICANT: Kuhnloss, Stuart A.
 : APPLICANT: Rostock, Paul R., Jr.
 : APPLICANT: Sutton, Kimberly L.
 : TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
 :
 : NUMBER OF SEQUENCES: 15
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: THOMAS G. PLANT 1501
 : STREET: LILLY CORPORATE CENTER
 : CITY: INDIANAPOLIS
 : STATE: IN
 : COUNTRY: USA
 : ZIP: 46285
 :
 : COMPUTER READABLE FORM:

```

1 MEDIUM TYPE: Floppy disk
2 COMPUTER: IBM Compatible
3 OPERATING SYSTEM: MS-DOS
4 SOFTWARE: ASCII(DOS) text only
5
6 CURRENT APPLICATION DATA:
7 APPLICATION NUMBER: US/08/804,227C
8 FILING DATE: February 21, 1997
9
10 CLASSIFICATION: 435
11
12 ATTORNEY/AGENT INFORMATION:
13 NAME: Plant, Thomas, G.
14 REGISTRATION NUMBER: 35,784
15 REFERENCE/DOCKET NUMBER: X-8231
16 TELECOMMUNICATION INFORMATION:
17 TELEPHONE: 317-276-2459
18
19 INFORMATION FOR SEQ ID NO: 1:
20 SEQUENCE CHARACTERISTICS:
21 LENGTH: 43280 base pairs
22 TYPE: nucleic acid
23 STRANDEDNESS: single
24 TOPOLOGY: linear
25
26 MOLECULE TYPE: DNA (genomic)
27
28 FEATURE:
29 NAME/KEY: CDS
30 LOCATION: 816..14234
31 FEATURE:
32 NAME/KEY: CDS
33 LOCATION: 14351..19945
34 FEATURE:
35 NAME/KEY: CDS
36 LOCATION: 20010..31199
37 FEATURE:
38 NAME/KEY: CDS
39 LOCATION: 31232..36067
40 FEATURE:
41 NAME/KEY: CDS
42 LOCATION: 36249..41774
43
44 US-08-804-227C-1

```

Query Match	7.6%	Score 75.8	DB 2	Length 43280
Best Local Similarity	43.0%	Pred. 4.9e-06		
Matches 371	Conservative	0	Mismatches 492	Indels 0
				Gaps 0
QY 35	gcccacgcgagggaagatctctgcacagtttcgcagacagccctgtctgtccagggucatcgcg	94		
Db 41236	gccttgccttcgtgacacatagccttcggggggccggggagctgcttccggcagaaacggc	41177		
QY 95	gaagccttcacgcgcgagagatcgagctgtctcggggttcgcgggcgatgaagcgccgcgcg	154		
Db 41176	ggcttcggcgacgagccggccatcggggcccggccggagagctccgcacgccccttcggcgaag	41117		
QY 155	atgcgcgggcgcgggaagcacacatgtcggagcccacactcggccgggtcttcgcgcgggttg	214		
Db 41116	gcgggaactctccacccggggccggagggccacccggcgtccggcccaaccacgacgacgggg	41057		
QY 215	cggacgaactctcgttagccttcagtcgcgcgggcgcagaggaagcgaacgaacctgcga	274		
Db 41056	agcccggtgtccttcggccgctgcgcgacacgacacagagggccaccggctgcgggtccgc	40997		
QY 275	cggcccttcgcgcgaattcacgcgcgtgtgcacgcgcgcgcgaactcttcacagcgtctcgttagc	334		
Db 40996	gtccggaaaggcggggcggaacccgggaagcggggcgagagatgttcggcattttccggggc	40937		
QY 335	aggccgcgcgttcagcgcgcgaacagcgaagaacgcgcgaagccgttagccttcgcgcgcgcg	394		
Db 40936	cgccgcgtgcgacacgccttcctccgtagagggccggcagaccctttccggacatccggccgc	40877		
QY 395	agcaccctcagctcgcgcgttcgcggagccgcgaagtgtacaggaactgtgtggaagccatg	454		
Db 40876	gcggcaccgggtcccgggccgcccggggtccgtgtcccggtccggtggcagacagtttcggcc	40817		
QY 455	cccaggaagtgacgcgcgcgcgtcttcctcctcgcgcgcgcgacgtgtgcacgaacgcgcgaag	514		

```

Db 40816 AGGTGGGCGGCGAGCGCGCGGGGTCTCCCGTCGAGAGCAGATCGGGGGCAGGTCTG 40757
QY 515 ttccagagagcgagcagtagagacattcgccgtctgagcagagagctgcatatgagctccac 574
Db 40756 AGGCGCGGTGGCCAGAGGAGTCTCCGCGCGGCGGACCGGTCAGGTGGCCAGGAGTGTGAAGCCC 40697
QY 575 acctgtcccaagcgagcgagcagctgagatgtgtgcatctgtgtagaaagctgagtgctg 634
Db 40696 ATGGCGGGGAGAACGTTGCGCCCTCGGCGACCTCGGCGGAGTGTGAGAGCCGACGCGCGG 40637
QY 635 accgcccagctgtgcagcgagcattccctgcagagagagagatgtgtgtgtgtgtgtgtgt 634
Db 40636 GCGGCTCGGAGCTGCGACCACTGCGCCAGCGCGGTGGGAGGCTGTGGGCGGCGGAGGCG 40577
QY 695 ctgtcgttagcagcgctgcagcattccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 754
Db 40576 GTTAGGTTCCGCGCCGACGAGGTGGGTCTCCCTCGGGGGTCTCCGCGCGCGCGCGCGCG 40517
QY 755 cgcgcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 814
Db 40516 CCGGCGGTCCGAGCGCCCGGATCAGAGGCTGTGGCGCGGAGGCGGTGAGGCGCTCGCGCG 40457
QY 815 agcgcgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 874
Db 40456 AAGCGCTCCAGTGGGTGTGGGTGAGCGGTGGGAGGCTTCCCGGAGGCGATGGCGGTG 40397
QY 875 tccatcagagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 897
Db 40396 GCCATCGCCGAGGAGCGGAGAC 40374

```

RESULT 9
US-08-042-747A-7/c
Sequence 7, Application US/08042747A
Patent No. 5487969

GENERAL INFORMATION:

APPLICANT: Eberle, Richard
APPLICANT: Black, Darla
APPLICANT: Scinciarillo, Franco
APPLICANT: Hilliard, Julia K.
TITLE OF INVENTION: Cloning and Amplification of Monkey B
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cox & Smith Incorporated
STREET: 112 East Pecan Street, Suite 2000
CITY: San Antonio
STATE: Texas
COUNTRY: USA
ZIP: 78205

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/042,747A
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Haymond, W. Bradley
REGISTRATION NUMBER: 35186
REFERENCE/DOCKET NUMBER: S-0072.179
TELECOMMUNICATION INFORMATION:
TELEPHONE: 210-554-5500
TELEFAX: 210-226-8395
TELEX: 767609

INFORMATION FOR SEQ. ID NO. 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 2943 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 87..2744
US-08-042-747A-7

Query Match 7.2%; Score 71.8; DB 1; Length 2943;
Best local Similarity 44.8%; Pred. No. 2.5e-05;
Matches 360; Conservative 0; Mismatches 437; Indels 6; Gaps 2;

```

QY 106 cgcgagatcgaagctgctcggagggttcggagccgatgaagcgagcgagcgagcgagcgag 165
Db 1535 CGTGCTTGTATCGCTCCACCGCTGGGGGGTCCACGAGGAGGCGCCGGGGTGGCCCTC 1476
QY 166 ggaacagaccatagtagagagagagagagagagagagagagagagagagagagagagagag 222
Db 1475 CCGGGGATCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1416
QY 223 ctctcgttagagctcgtatcgccgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 282
Db 1415 CTCACAGAGCGCTTGTCTGAGAGAGGGGCTGTGACGAGATGAGAGAACCCCGCTGGCCAG 1356
QY 283 cgcgcgacttaccgcgagtgctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 342
Db 1355 GTAGTACTGACCTGGCGCCACCTTCACGTGCGCTCCGTTGTACGCGCGGAGAGATGGG 1296
QY 343 gtgcagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 402
Db 1295 GTCCAGGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1236
QY 403 cagctcgcgctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 462
Db 1235 GTACTCGGTGCGGTGTGGCGGTGAAGGTGTGAGAGAGGCGCGCGCGCGCGCGCGCGCGAG 1176
QY 463 gtgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 519
Db 1175 GGGGCGGTACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1116
QY 520 cgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 579
Db 1115 CGAGGGGCGCTTGGGGGCCAGTCCAGTCCAGCCAGCGTGAAGTGGGGGTGTGACGAGGTT 1056
QY 580 gtcccaagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 639
Db 1055 GCGCGTACACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 996
QY 640 cagcctgcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 699
Db 995 GACCTCGCGGAAAGCGGTCCCGGGGCGTGAAGCGCTGCTCCCGTGGGACCGCGCGGTA 936
QY 700 gtgcagcgagctgcgtatcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 759
Db 935 GCGGTGAAGCGGAGCATGATGACAAAGTCCCGGTGCGCAGCAGCACTCGTGTACGG 876
QY 760 tccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 819
Db 875 GTACACCGAGCGGCGCTCCACCTCTCGACGATACAGTTACCCGTGTGCGCGTGAAGCGGTG 816
QY 820 ccagcgcgtatcgtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 879
Db 815 GAAGGCGCTTCACCCGCGCGGGGTTGTACTTGAAGTGGGTGTGTCACCGCGCGGCTGGT 756
QY 880 cagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 902
Db 755 GCGGCTCGCGCGCTTGGCGCGGT 733

```

RESULT 10
5212296-8/c
Patent No. 5212296
APPLICANT: DEAN, CAROLINE; HARDER, PATRICIA A.; LETO, KENNETH
J.; O'KEEFE, DANIEL P.; OMER, CHARLES A.; ROMESSER, JAMES A.


```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: System Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806, 0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
US-08-510-646B-1

```

query Match	7.1%;	Score 71.2;	DB 3;	Length 5392;
Best Local Similarity	45.4%;	Pred. No. 3.2e-05;		
Matches 305;	Conservative 0;	Mismatches 358;	Indels 9;	Gaps 1;

```

OY      887 agecgcgctcg 898
        |         ||||| 1
Db      2687 ATTCCTTCGTGG 2676

RESULT 12
US-08-510-646B-1/c
: Sequence 1, Application US/08510646B
: Patent No. 6077699
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanchet, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Creely-Iagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

[illegible]

